The Annals of Probability
1996, Vol. 24, No. 2, 698–742

A COUNTABLE REPRESENTATION OF THE FLEMING–VIOT MEASURE-VALUED DIFFUSION

BY PETER DONNELLY[1] AND THOMAS G. KURTZ[2]

Queen Mary and Westfield College and University of Wisconsin–Madison

The Fleming–Viot measure-valued diffusion arises as the infinite population limit of various discrete genetic models with general type space. The paper gives a countable construction of the process as the empirical measure carried by a certain interactive particle system. This explicit representation facilitates the study of various properties of the Fleming–Viot process. The construction also carries versions of the familiar genealogical processes from population genetics, in particular, Kingman's coalescent, thus unifying the genealogical and measure-valued approaches to the subject.

Introduction. The approach to infinite population genetics models taken by Fleming and Viot (1979) views gene frequencies as probability distributions on some space E of genetic "types." [See Ethier and Kurtz (1993) for a recent survey of Fleming—Viot processes.] It is useful to let this space of types be very general [see, e.g., Ethier and Griffiths (1987, 1990)], so we will allow E to be any complete, separable metric space. The Fleming–Viot process is a Markov process with sample paths in $\mathscr{P}(E)$, the space of probability measures on E. Our primary goal in the present paper is to introduce an E^∞-valued process $X = (X_1, X_2, \dots)$ whose infinite empirical measure

$$Z(t) \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k(t)} \tag{I.1}$$

is a Fleming–Viot process. A process of this type appears implicitly in the work of Dawson and Hochberg (1982), who used the process to study the support properties of the Fleming–Viot process with Brownian mutation. Here we will see that the desired process is Markov with a generator of a

---

Received March 1994; revised March 1995.

[1] Research supported by SERC Advanced Fellowship B/AF/1255 and Grants GR/F 98727 and GR/G 11101. Current address: Departments of Statistics, and Ecology and Evolution, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637.

[2] Supported by SERC Visiting Fellowship GR/G27485 and NSF Grants DMS-88-03284 and DMS-92-04866.

AMS 1991 *subject classifications*. 60J25, 60K35, 60J70, 60J80, 92D10.

Key words and phrases. Fleming–Viot process, measure-valued diffusion, exchangeability, coupling, sample-path properties, ergodicity, genealogical processes, the coalescent.

simple form and that an explicit construction of the process can be given in terms of a collection of Poisson processes and E-valued Markov processes. This explicit construction allows us to derive a variety of properties of the Fleming–Viot process under very general assumptions on the mutation process.

Along with recent interest in measure-valued processes in genetics, there has been a complementary focus on the genealogical structure induced by genetics models. Genealogical techniques have provided powerful tools for studying such models. The particle model we construct not only carries the Fleming–Viot process, but it also contains within it the familiar genealogical processes, including Kingman's coalescent. The construction in this paper thus provides a unification of these previously (formally) separate approaches to infinite population genetics models. In addition, it carries various urn models which have recently emerged [e.g., Hoppe (1987) and Ethier and Griffiths (1987)] as convenient devices for constructing probability distributions of interest. One practical consequence of this unification is that results or calculations derived from one approach immediately apply to the other, without the need for separate proofs.

In the next section we introduce the simplest model for demography in population genetics, the so-called Moran model. The Fleming–Viot process can be obtained as the limit of the empirical measure processes associated with a suitably scaled sequence of Moran models. Certain symmetries in the resulting generators motivate our "particle model" construction of the process.

The discrete construction of the process, as the empirical measure associated with a certain infinite particle system, is introduced in Section 2. A key result, Lemma 2.1, shows that for each n, the first n particles in this system may be coupled with a Moran model of size n, in such a way that the empirical measures of each process are the same. It follows that the marginal distributions of the E^{∞}-valued particle process are exchangeable and hence that the associated empirical measure exists (as the de Finetti representing measure). As it also arises as the limit of the empirical measures associated with the Moran models, it should correspond to the Fleming–Viot measure. The main result of the section, Theorem 2.4, extends this "finite dimensional" convergence to show that the empirical measure associated with the particle system exists as a process and that it is a version of the appropriate Fleming–Viot process.

Section 3 is concerned with the genealogical structure of the particle process. We show that n-coalescents, for each n, and the (infinite) coalescent arise by tracing "ancestry" in the particle system. One can reconstruct the types of the particles, and hence sampling distributions (or moment measures) and in fact the empirical measure itself, by tracing forward through the genealogy from the common ancestor, superimposing the effects of the mutation processes. This gives rise to (generalizations of) various urn-type models which have recently proved to be valuable tools for studying these distributions.

The final two sections illustrate the use of the particle model, and the associated genealogical structure, in deriving various results about the Fleming–Viot process. Section 4 is concerned with ergodicity and estimates of distance from stationarity and Section 5 with sample-path properties of the process.

1. The Moran model. We consider a sequence $\{V^n\}$ of Moran models [Moran (1958)] with mutation probabilities γ_n and mutation distributions $\eta_n(x,\cdot)\in\mathscr{P}(E)$, $x\in E$. That is, for $n=2,3\dots$, V^n is a continuous time, E^n-valued Markov process representing the types of the individuals in a population of size n which evolves as follows. With intensity $\frac{1}{2}$, the jth individual in the population dies and is replaced by an offspring of the ith individual $(i\neq j)$. With probability $1-\gamma_n$ the offspring has the same type y_i as the parent, and with probability γ_n the type of the offspring is randomly determined according to the distribution $\eta_n(y_i,\cdot)$. Define

$$\begin{aligned} Q^n_{ij}f(y_1,\dots,y_n) \\ = \gamma_n\int f(y_1,\dots,y_{j-1},z,y_{j+1},\dots,y_n)\eta_n(y_i,dz) \\ +(1-\gamma_n)f\big(\theta_{ij}(y_1,\dots,y_n)\big), \end{aligned} \tag{1.1}$$

where $\theta_{ij}(y)\in E^n$ is obtained from y by replacing y_j by y_i. Then the generator of the nth Moran model is

$$C_n f=\frac{1}{2}\sum_{1\le i\neq j\le n}\big(Q^n_{ij}f-f\big) \tag{1.2}$$

for $f\in B(E^n)$ [where $B(S)$ denotes the bounded Borel measurable functions on a metric space S].

We want to derive the Fleming–Viot process as the limit of the sequence of Moran models as the population size $n\to\infty$. The critical hypothesis for convergence can be stated in terms of the sequence of operators defined by

$$B^n g(x)=\frac{n\gamma_n}{2}\int(g(z)-g(x))\eta_n(x,dz) \tag{1.3}$$

for $g\in B(E)$. Note that B^n is the generator of a pure jump Markov process on E, and we assume that B^n converges to the generator B of an E-valued Markov process and that the corresponding sequence of jump Markov processes converges in distribution to the limiting process in the Skorohod topology. Note that the limiting process need not be pure jump. (In the original work of Fleming and Viot, the limiting process was Brownian motion.)

For $y\in E^n$, define $\mu_y\in\mathscr{P}(E)$ by $(1/n)\sum\delta_{y_i}$, and for $g\in B(E^m)$, $m<n$, define $\Gamma g\in B(E^n)$ by defining $g_{i_1\cdots i_m}(y)\equiv g(y_{i_1},\dots,y_{i_m})$ and

$$\Gamma g(y)\equiv\frac{1}{n^m}\sum_{1\le i_1\neq\cdots\neq i_m\le n}g_{i_1\cdots i_m}(y). \tag{1.4}$$

In the following computation, for a vector x and an element u, $\theta_k(x|u)$ is the vector of the same length obtained by replacing the kth component of x by u [e.g., $\theta_{ij}(x) = \theta_j(x|x_i)$]; for $f \in B(E^m)$, $\Phi_{ij} f(x) \equiv f(\theta_{ij}(x))$ and

$$B_k^n f(x) = \frac{n\gamma_n}{2} \int \big(f(\theta_k(x|z)) - f(x)\big)\eta_n(x_k, dz). \tag{1.5}$$

Note that B_k^n is just B^n applied to the kth variable in f. Note also that if $f(y)$ does not depend either on y_i or y_j, then $Q_{ij} f(y) = f(y)$. It follows that

$$\begin{aligned}
C_n \Gamma g &= \frac{1}{n^m} \sum_{1 \le i_1 \ne \cdots \ne i_m \le n} C_n g_{i_1 \cdots i_m}(y) \\
&= \frac{1}{2n^m} \sum_{1 \le i_1 \ne \cdots \ne i_m \le n} \sum_{k=1}^{m} \Bigg(\sum_{j \notin \{i_1, \dots, i_m\}} \big(Q_{j i_k}^n g_{i_1 \cdots i_m} - g_{i_1 \cdots i_m}\big) \\
&\qquad + \sum_{l \ne k} \big(Q_{i_l i_k}^n g_{i_1 \cdots i_m} - g_{i_1 \cdots i_m}\big)\Bigg) \\
(1.6) \quad &= \frac{1}{2n^m} \sum_{1 \le i_1 \ne \cdots \ne i_m \le n} \sum_{k=1}^{m} \Bigg(\sum_{j \notin \{i_1, \dots, i_m\}} \bigg(\frac{2}{n} B_k^n g_{\theta_k(i_1, \dots, i_m | j)} \\
&\qquad + \big(g_{\theta_k(i_1, \dots, i_m | j)} - g_{i_1 \cdots i_m}\big)\bigg) \\
&\qquad + \sum_{l \ne k} \bigg(\frac{2}{n} B_k^n g_{\theta_{lk}(i_1, \dots, i_m)} + (1 - \gamma_n)\big(g_{\theta_{lk}(i_1, \dots, i_m)} - g_{i_1 \cdots i_m}\big)\bigg)\Bigg) \\
&= \sum_{k=1}^{m} \frac{n-m}{n} \Gamma B_k^n g + \frac{1}{2} \sum_{1 \le l \ne k \le m} \bigg(\frac{2}{n} \Gamma \Phi_{lk} B_k^n g + (1 - \gamma_n)\Gamma(\Phi_{lk} g - g)\bigg),
\end{aligned}$$

where in the last step we have used the fact that

$$\sum_{1 \le i_1 \ne \cdots \ne i_m \ne j \le n} g_{i_1 \cdots i_m} = \sum_{1 \le i_1 \ne \cdots \ne i_m \ne j \le n} g_{\theta_k(i_1, \dots, i_m | j)} = (n - m)\Gamma g.$$

For $g \in B(E^m)$ and $\mu \in \mathscr{P}(E)$, define $G(\mu) \equiv \langle g, \mu^m \rangle$ to be the integral of g against the m-fold product measure for μ. As $n \to \infty$,

$$\sup_{y \in E^n} |\Gamma g(y) - \langle g, \mu_y^m \rangle| \to 0,$$

and if $\gamma_n \to 0$ and for each k, $B_k^n g \to B_k g$ uniformly on E^m, then $\sup_{y \in E^n} |C_n \Gamma g(y) - \mathbb{A} G(\mu_y)| \to 0$, where

$$\mathbb{A} G(\mu) = \sum_{k=1}^{m} \langle B_k g, \mu^m \rangle + \sum_{1 \le k \ne l \le m} \big(\langle \Phi_{lk} g, \mu^{m-1} \rangle - \langle g, \mu^m \rangle\big). \tag{1.7}$$

For $V^n = (V_1^n, \dots, V_n^n)$ defined above, let Z_n be defined by

$$Z_n(t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{V_i^n(t)}. \tag{1.8}$$

Then under appropriate conditions on the convergence of B_k^n to B_k and the assumption that $\{Z_n(0)\}$ converges in distribution, the above calculation implies that Z_n converges in distribution in the Skorohod space $D_{\mathscr{P}(E)}[0,\infty)$ to a $\mathscr{P}(E)$-valued Markov process Z whose generator is given by (1.7). [See Ethier and Kurtz (1986), Chapter 10.]

To be precise about the conditions we will place on the mutation process, let (E, r) be a complete, separable metric space, and let $P(t, x, G)$ be a transition function corresponding to a Markov process with sample paths in $D_E[0,\infty)$. [In particular, $P(t, x, G)$ is assumed to be jointly measurable in (t, x) for each $G \in \mathscr{B}(E)$.] Let P_x denote the distribution on $D_E[0,\infty)$ of the version of the process starting from x, and note that $x \to P_x$ is a Borel measurable mapping of E into $\mathscr{P}(D_E[0,\infty))$. This process will be the mutation process for our model. Define the corresponding semigroup on $B(E)$ by

$$T(t)f(x) = \int_E f(y)P(t, x, dy). \tag{1.9}$$

Let $\tilde{L}_0 = \{f \in B(E): \text{bp-lim}_{t\to 0} T(t)f = f\}$. (bp-lim will denote the bounded, pointwise limit.) The assumed right continuity of the corresponding processes implies that $\overline{C}(E) \subset \tilde{L}_0$ [where $\overline{C}(E)$ denotes the bounded continuous functions on E], which in turn implies that $\tilde{L}_0$ is bp-dense in $B(E)$, that is, $B(E)$ is the smallest collection of functions that contains $\tilde{L}_0$ and is closed under bounded-pointwise convergence. Define

$$Bf \equiv \underset{t\to 0}{\text{bp-lim}} \frac{T(t)f - f}{t} \tag{1.10}$$

when the limit exists. [Note that this is the "weak infinitesimal operator" for $\{T(t)\}$ as defined in Dynkin (1965). We could work with the "full generator" as defined in Ethier and Kurtz (1986), but the weak infinitesimal operator is notationally simpler.] B will be called the *mutation operator* for the process. We will need the following consequence of these conditions for the mutation operator.

LEMMA 1.1. *There exists a countable subset $D \subset \mathscr{D}(B)$, the domain of B, that is separating for $\mathscr{P}(E)$ in the sense that for $\mu, \nu \in \mathscr{P}(E)$, $\int f\,d\mu = \int f\,d\nu$ for all $f \in D$ implies that $\mu = \nu$.*

PROOF. By the separability of E, there exists a countable separating subset $\hat{D} \subset \overline{C}(E)$. Let D contain all functions of the form

$$f_\lambda = \int_0^\infty e^{-\lambda t}T(t)g\,dt \tag{1.11}$$

for $\lambda \in (0,\infty) \cap \mathbb{Q}$, $g \in \hat{D}$. Since $\text{bp-lim}_{\lambda\to\infty} \lambda f_\lambda = g$, it follows that D is separating. □

Define $\{T(t)\}$ on $B(E^\infty)$ to be the semigroup corresponding to the E^∞-valued Markov process $X = (X_1, X_2, \dots)$ each of whose components is Markov with transition function $P(t, x, G)$, with $\{X_k\}$ conditionally independent given $\{X_k(0)\}$. In particular, for $f \in B(E^m)$,

$$T(t)f(x_1, \dots, x_m) = \int_E \cdots \int_E f(y_1, \dots, y_m) P(t, x_1, dy_1) \cdots P(t, x_m, dy_m). \tag{1.12}$$

Let B also denote the weak infinitesimal operator for $\{T(t)\}$ on $B(E^\infty)$. We will write

$$Bf(x_1, \dots, x_m) = \sum_{k=1}^{m} B_k f(x_1, \dots, x_m), \tag{1.13}$$

where B_k is B applied to f as a function of the kth variable alone, when the right side is defined.

For $1 \le i < j \le m$, define the *replacement operators* Φ_{ij}: $B(E^m) \to B(E^{m-1})$ by letting $\Phi_{ij} f$ be the function obtained from f by replacing x_j by x_i and renumbering the variables [e.g., for $f(x_1, x_2, x_3) \in B(E^3)$, $\Phi_{12} f(x_1, x_2) = f(x_1, x_1, x_2)$ and $\Phi_{23} f(x_1, x_2) = f(x_1, x_2, x_2)$].

For $f \in B(E^m)$, define $F_f \in B(\mathscr{P}(E))$ by $F_f(\mu) = \langle f, \mu^m \rangle$, where μ^m denotes the m-fold product measure of μ, and let $\mathbb{L}^m(E) \subset B(\mathscr{P}(E))$ be the collection of functions of this form. Note that $\mathbb{L}^m(E)$ is a linear subspace of $B(\mathscr{P}(E))$. For $f \in \mathscr{D}(B) \cap B(E^m)$, the generator of the neutral Fleming–Viot process $\mathbb{A}$ is given by

$$\mathbb{A} F_f(\mu) = \langle Bf, \mu^m \rangle + \sum_{1 \le i < j \le m} \bigl(\langle \Phi_{ij} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle \bigr). \tag{1.14}$$

Since for $f \neq g$, one may still have $F_f = F_g$, it is not clear from this definition that $\mathscr{A}$ is single-valued, nor is it clear that $\mathbb{A}$ is dissipative (i.e., satisfies $\|\lambda F - \mathbb{A} F\| \ge \lambda \|F\|$, $\lambda > 0$). Consequently, we should really define $\mathbb{A}$ as the set of ordered pairs $\{(F_f, G_f): f \in B(E^m),\ m = 1, 2, \dots\}$ with G_f defined by the right side of (1.14). The martingale problem for multivalued generators is developed in Ethier and Kurtz (1986) and that theory can be applied here. Consequently, we do not need to verify single-valuedness here. In Section 2, we will prove the existence of solutions of the martingale problem for $\mathbb{A}$ for every initial distribution. This existence implies that $\mathbb{A}$ is, in fact, dissipative [Ethier and Kurtz (1986), Proposition 4.3.5]. Under some conditions, dissipativity in turn implies single-valuedness. If, for example, E is compact and $D_1 = \{f \in \mathscr{D}(B) \cap C(E)\colon f, Bf \in C(E)\}$ is dense in $C(E)$, then $D_m = \{f \in \mathscr{D}(B) \cap C(E^m)\colon Bf \in C(E^m)\}$ is dense in $C(E^m)$ and $\mathbb{D} = \{F_f\colon f \in D_m, m = 1, 2, \dots\}$ is dense in $C(\mathscr{P}(E))$. It follows that $\mathbb{A}$ restricted to $\mathbb{D}$ is single-valued. [See Ethier and Kurtz (1986), Lemma 1.4.2.] If A defined in (1.15) below is a restriction of the weak infinitesimal operator for $\{S(t)\}$ in (1.18) [which will be the case if $\mathscr{D}(B) \cap B(E^m) \subset \overline{C}(E)$], then it follows from (1.18) that $\mathbb{A}$ is a restriction of the weak infinitesimal operator. In general, we are not aware of a direct proof of single-valuedness of $\mathbb{A}$. We will, however, continue to use the simpler "single-valued" notation $\mathbb{A}F_f$ with the understanding that the martingale property must hold for all "values" of $\mathbb{A}F_f$.

We are interested in the relationship between this Fleming–Viot process and the E^∞-valued process whose generator for $f \in \mathscr{D}(B) \cap B(E^m)$ is given by

$$\begin{aligned} Af(x_1, \dots, x_m) &= Bf(x_1, \dots, x_m) \\ &\quad + \sum_{1 \le i < j \le m} \bigl(f(\theta_{ij}(x_1, \dots, x_m)) - f(x_1, \dots, x_m)\bigr), \end{aligned} \tag{1.15}$$

where $\theta_{ij}(x_1, \dots, x_m)$ denotes the element of E^m obtained by replacing x_j by x_i in $(x_1, \dots, x_m)$. The jth component of this process will evolve as a Markov process with generator B independently (conditioned on its initial position) of the other components for a period of time exponentially distributed with parameter $j - 1$. It then "looks down" at a component chosen at random from among the first $j - 1$, assumes the value of that component and then evolves independently for another exponentially distributed length of time. Note that the restriction of A to $B(E^m)$ is a bounded perturbation of the restriction of B to $B(E^m)$, so existence of this process (as a solution of the martingale problem for A) follows from Ethier and Kurtz [(1986), Proposition 4.10.2]. Furthermore, $\mathscr{R}(\lambda - B) \cap B(E^m)$ is bounded-pointwise dense in $B(E^m)$ and the fact that $A|_{B(E^m)}$ is a bounded perturbation of $B|_{B(E^m)}$ implies $\mathscr{R}(\lambda - A) \cap B(E^m)$ is bounded-pointwise dense in $B(E^m)$. It follows that $\mathscr{R}(\lambda - A)$ is bounded-pointwise dense in $B(E^\infty)$. In particular, the full generator $\hat{A}$ [see Ethier and Kurtz (1986), Section 1.5] is the bounded-pointwise closure in $B(E^\infty) \times B(E^\infty)$ of $\{(f, Af)\colon f \in \mathscr{D}(B)\}$. Uniqueness of the process (that is, uniqueness for solutions of the martingale problems for A) follows from Theorem 4.4.1 of Ethier and Kurtz (1986).

To see that the process with generator A is closely related to the Fleming–Viot process with generator $\mathbb{A}$, observe that if $f \in \mathscr{D}(B) \cap B(E^m)$ and $F(\mu) = \langle f, \mu^m \rangle$, then

$$\mathbb{A}F(\mu) = \langle Af, \mu^m \rangle \tag{1.16}$$

and for $g \in B(E^m)$ and $G(\mu) = \langle g, \mu^m \rangle$

$$(\lambda - \hat{\mathbb{A}})^{-1} G(\mu) = \langle (\lambda - \hat{A})^{-1} g, \mu^m \rangle. \tag{1.17}$$

It follows that

$$\mathbb{S}(t)G(\mu) = \langle S(t)g, \mu^m \rangle, \tag{1.18}$$

where $\{\mathbb{S}(t)\}$ is the semigroup corresponding to the Fleming–Viot process and $\{S(t)\}$ is the semigroup corresponding to A. In Section 2, we will see that the relationship between the Fleming–Viot process and the particle process is much more than a functional analytic identity.

2. Construction of the particle model. In order to be able to make an explicit construction of the desired particle model, we assume that there is a probability space $(\Omega_0, \mathscr{F}_0, P_0)$ and a $\mathscr{B}(E) \times \mathscr{B}[0, \infty) \times \mathscr{F}_0$-measurable mapping $U\colon E \times [0, \infty) \times \Omega_0 \to E$ such that $U(x, 0, \omega) = x$ for all x and ω, and that for each x, $U(x,\cdot,\cdot)$ is a Markov process with transition function $P(t,x,G)$ and sample paths in $D_E[0,\infty)$. For example, existence of such a U follows from the measurability of the mapping $x \to P_x$ and the construction of Blackwell and Dubins (1983). In fact, if the mapping $x \to P_x$ is continuous as a mapping from E into $D_E[0,\infty)$ (which it typically will be), the Blackwell and Dubins construction ensures that U is almost surely continuous at each x. That is, for each x, $P\{\omega\colon \lim_{y\to x} U(Y,\cdot,\omega) = U(x,\cdot,\omega)\} = 1$, where convergence is in the Skorohod topology. Alternatively, for diffusion processes that can be obtained as unique, strong solutions of Itô equations, $U(x,\cdot,\cdot)$ can simply be the solution of the Itô equation with initial value x.

Let $\{U_{ijk}, 1 \le i < j, 1 \le k < \infty\}$ and $\{U_{i0}, i \ge 1\}$ be independent realizations of U (thought of as an E-valued stochastic process with index set $E \times [0,\infty)$), let $\{N_{ij}, 1 \le i < j\}$ be independent, unit rate Poisson processes, independent of $\{U_{ijk}\}$ and let τ_{ijk} denote the kth jump time of N_{ij}. Let $\{X_i(0), i \ge 1\}$ be an exchangeable sequence of E-valued random variables, independent of $\{U_{ijk}\}$, $\{U_{i0}\}$ and $\{N_{ij}\}$. Define $\gamma_{ijk} = \min\{\tau_{i'jk'}, i' < j\colon \tau_{i'jk'} > \tau_{ijk}\}$; that is, γ_{ijk} is the first jump time of $N_j \equiv \sum_{i<j} N_{ij}$ after τ_{ijk}, and define $\gamma_{j0} = \min\{\tau_{ij1}\colon i < j\}$. Finally, define

$$X_j(t) = U_{j0}\big(X_j(0), t\big), \qquad 0 \le t < \gamma_{j0}, \tag{2.1}$$

and

$$X_j(t) = U_{ijk}\big(X_i(\tau_{ijk}), t - \tau_{ijk}\big), \qquad \tau_{ijk} \le t < \gamma_{ijk}. \tag{2.2}$$

Note that between the jump times of the Poisson processes, the X_j behave as independent Markov processes with transition function $P(t,x,G)$ and that at the jump times of N_{ij}, X_j "looks down" at X_i, assumes the value of X_i at the jump time, and then evolves independently according to $P(t,x,G)$ until the next jump time of N_j. We see that X is the desired process with generator (1.15).

To better understand the properties of the E^∞-valued process $X = (X_1, X_2, \ldots)$, we want to compare it to a sequence of Moran-type models. (These models differ from standard Moran models in that mutation may occur any time during the life of an individual rather than only at birth. Both types of Moran models, appropriately rescaled, converge to the Fleming–Viot process.) Let $\{\tilde N_{ij}, 1 \le i \ne j \le m\}$ be independent Poisson processes with intensity $\frac{1}{2}$, let $\{\tilde U_{j0}, 1 \le j \le m\}$ and $\{\tilde U_{ijk}, 1 \le i \ne j \le m, k \ge 1\}$ be independent copies of U and let $Y_1^m(0), \ldots, Y_m^m(0)$ be exchangeable E-valued random variables, independent of the other processes. Let $\tilde\tau_{ijk}$ be the kth jump time of $\tilde N_{ij}$, $\tilde\gamma_{ijk} = \min\{\tilde\tau_{i'jk'}\colon i' \ne j, \tilde\tau_{i'jk'} > \tilde\tau_{ijk}\}$ and let $\tilde\gamma_{j0}$ be the first jump time of $\tilde N_j \equiv \sum_{i \ne j} \tilde N_{ij}$. Define

$$Y_j^m(t) = \begin{cases} \tilde U_{j0}\big(Y_j^m(0), t\big), & 0 \le t < \tilde\gamma_{j0}, \\ \tilde U\big(Y_i^m(\tilde\tau_{ijk}), t - \tilde\tau_{ijk}\big), & \tilde\tau_{ijk} \le t < \tilde\gamma_{ijk}. \end{cases} \tag{2.3}$$

Note that the components of $Y^m = (Y_1^m, \dots, Y_m^m)$ behave as independent Markov processes with transition function $P(t, x, G)$ except that at the jump times of $\tilde{N}_{ij}$, Y_j^m is set equal to Y_i^m. In particular, with B as above, the generator for Y^m is given by

$$
\begin{aligned}
&A_m f(y_1, \dots, y_m) \\
&\quad = Bf(y_1, \dots, y_m) \\
&\qquad + \tfrac{1}{2} \sum_{1 \le i \ne j \le m} \Big(f\big(\theta_{ij}(y_1, \dots, y_m)\big) - f(y_1, \dots, y_m)\Big),
\end{aligned} \tag{2.4}
$$

where, as before, $\theta_{ij}(y_1, \dots, y_m)$ is obtained from $(y_1, \dots, y_m)$ by replacing y_j by y_i.

We now want to construct $Y^m = (Y_1^m, \dots, Y_m^m)$ in a particular way on the same sample space as X. Let S_m denote the collection of permutations of $(1, \dots, m)$ which we write as ordered m-tuples $s = (s_1, \dots, s_m)$. Let $\pi_{ij}\colon S_m \to S_m$ denote the mapping such that $\pi_{ij} s$ is obtained from s by interchanging s_i and s_j and let $\{M_{ijk}\colon 1 \le i \ne j \le m, k \ge 1\}$ be independent random mappings $M_{ijk}\colon S_m \to S_m$ such that $P\{M_{ijk} s = s\} = P\{M_{ijk} s = \pi_{ij} s\} = \frac{1}{2}$. Define an S_m-valued process Σ^m and counting processes $\{\tilde{N}_{ij}, 1 \le i \ne j \le m\}$ as follows. Let $\Sigma^m(0)$ be uniformly distributed on S_m and independent of all other processes. Let

$$
\tilde{N}_{ij}(t) = \sum_{1 \le k < l \le m} \int_0^t I_{\{\Sigma_i^m(r-) = k,\ \Sigma_j^m(r-) = l\}}\, dN_{kl}(r) \tag{2.5}
$$

and let Σ^m be constant except for discontinuities determined by $\Sigma^m(\tilde{\tau}_{ijk}) = M_{ijk} \Sigma^m(\tilde{\tau}_{ijk} -)$, where $\tilde{\tau}_{ijk}$ is the kth jump time of $\tilde{N}_{ij}$, or more precisely, interpreting Σ^m as a $\mathbb{Z}^m$-valued process,

$$
\Sigma^m(t) = \sum_{1 \le i < j \le m} \int_0^t \Big(M_{ij\tilde{N}_{ij}(r-)+1} \Sigma^m(r -)\Big)\, d\tilde{N}_{ij}(r). \tag{2.6}
$$

Equations (2.5) and (2.6) determine the $\tilde{N}_{ij}$ and Σ^m recursively. In addition, define $\{\hat{N}_{ij}, 1 \le i \le m < j\}$ by

$$
\hat{N}_{ij}(t) = \sum_{k=1}^{m} \int_0^t I_{\{\Sigma_i^m(r-) = k\}}\, dN_{kj}(r) \tag{2.7}
$$

and let $\hat{\tau}_{ijk}$ denote the kth jump time of $\hat{N}_{ij}$. Note that for $j > m$,

$$
N_j = \sum_{1 \le i < j} N_{ij} = \sum_{1 \le i \le m} \hat{N}_{ij} + \sum_{m < i \le j} N_{ij}. \tag{2.8}
$$

LEMMA 2.1. (a) *The* $\{\tilde{N}_{ij}\}$ *and* $\{\hat{N}_{ij}\}$ *constructed above are independent Poisson processes; the* $\tilde{N}_{ij}$ *have intensity* $\frac{1}{2}$ *and the* $\hat{N}_{ij}$ *have intensity* 1, *and for each* $t \ge 0$, $\Sigma^m(t)$ *is independent of* $\mathscr{G}_t \equiv \sigma(\tilde{N}_{ij}(s), \hat{N}_{kl}(s)\colon s \le t, 1 \le i \ne j \le m, 1 \le k \le m < l)$.

(b) *Define* $Y_j^m(t) = X^m_{\Sigma_j^m(t)}$, $j = 1, \dots, m$. *Then* $Y^m = (Y_1^m, \dots, Y_m^m)$ *is Markov with generator* (2.4).

(c) *For* $i \le m < j$, *let* $\tilde{U}_{ijk} = U_{\alpha,j,\beta}$, *where* $\alpha = \Sigma_i^m(\hat{\tau}_{ijk}-)$ *and* $\beta = N_{\Sigma_i^m(\hat{\tau}_{ijk}-)j}(\hat{\tau}_{ijk})$. *Then for* $j > m$, X_j *satisfies*

$$
(2.9)\quad X_j(t) = \begin{cases} U_{j0}\big(X_j(0), t\big), & 0 \le t < \gamma_{j0}, \\ U_{ijk}\big(X_i(\tau_{ijk}), t - \tau_{ijk}\big), & \tau_{ijk} \le t < \gamma_{ijk}, m < i < j, \\ \tilde{U}_{ijk}\big(Y_i^m(\hat{\tau}_{ijk}), t - \hat{\tau}_{ijk}\big), & \hat{\tau}_{ijk} \le t < \gamma_{ijk}, i \le m, \end{cases}
$$

and $\Sigma^m(t)$ *is independent of* $\mathscr{H}_t^m = \sigma(Y^m(s), X_{m+1}(s), X_{m+2}, \dots, s \le t)$.

Proof. Let $\tilde{N}$ denote the array $\{\tilde{N}_{ij}\}$ and let $\hat{N}$ denote the array $\{\hat{N}_{ij}\}$. Then the process $(\Sigma^m, \tilde{N}, \hat{N})$ is a Markov process with generator

$$
\begin{aligned}
&Cf(s,u,v) \\
(2.10)\quad &= \sum_{1 \le i \ne j \le m} I_{\{s_i < s_j\}} \left(\frac{f(\pi_{ij}s, u + e_{ij}, v) + f(s, u + e_{ij}, v)}{2} - f(s,u,v) \right) \\
&\quad + \sum_{1 \le i \le m < j} \big(f(s, u, v + e_{ij}) - f(s,u,v) \big),
\end{aligned}
$$

where e_{ij} denotes the array with ijth element 1 and all other elements 0. Let Υ denote the uniform distribution on S_m. Then

$$
\begin{aligned}
&\int Cf(s,u,v)\Upsilon(ds) \\
(2.11)\quad &= \sum_{1 \le i \ne j \le m} \tfrac{1}{2} \left(\int f(s, u + e_{ij}, v)\Upsilon(ds) - \int f(s,u,v)\Upsilon(ds) \right) \\
&\quad + \sum_{1 \le i \le m < j} \left(\int f(s, u, v + e_{ij})\Upsilon(ds) - \int f(s,u,v)\Upsilon(ds) \right).
\end{aligned}
$$

The right side of (2.11) is the generator for an array of independent Poisson processes applied to the function $g(u,v) = \int f(s,u,v)\Upsilon(ds)$. If $(\tilde{L}, \hat{L})$ is such an array, for all bounded f,

$$
(2.12)\quad \int f\big(s, \tilde{L}(t), \hat{L}(t)\big)\Upsilon(ds) - \int_0^t \int Cf\big(s, \tilde{L}(r), \hat{L}(r)\big)\Upsilon(ds)\, dr
$$

is a martingale with respect to the filtration $\mathscr{F}_t^L = \sigma(\tilde{L}(r), \hat{L}(r): r \le t)$. Consequently, $(\Upsilon, \tilde{L}, \hat{L})$ is a solution of the filtered martingale problem for C in the sense of Kurtz and Ocone (1988). By the uniqueness of such solutions, it follows that $(\tilde{N}, \hat{N})$ is an array of independent Poisson processes, $\tilde{N}$ with parameter $\frac{1}{2}$ and $\hat{N}$ with parameter 1, and that the conditional distribution of $\Sigma^m(t)$ given $\mathscr{G}_t$ is Υ. In particular, $\Sigma^m(t)$ is independent of $\mathscr{G}_t$ and part (a) follows.

To prove part (b), for $1 \le i$, $j \le m$, define

$$
\begin{aligned}
&\tilde{U}_{j0} = U_{\alpha 0}, && \text{where } \alpha = \Sigma_j^m(0), \\
(2.13)\quad &\tilde{U}_{ijk} = U_{\alpha,\beta,\gamma}, && \text{where } \alpha = \Sigma_i^m\big(\tilde{\tau}_{ijk}-\big), \beta = \Sigma_j^m\big(\tilde{\tau}_{ijk}-\big), \\
& && \gamma = N_{\Sigma_i^m(\tilde{\tau}_{ijk}-), \Sigma_j^m(\tilde{\tau}_{ijk}-)}
\end{aligned}
$$

and $Y_j^m(0) = X_{\Sigma_j^m(0)}(0)$, $1 \le i \le m$. Since, by definition, $\{U_{j0}\}$, $\{U_{ijk}\}$ and $\{X_j(0)\}$ are independent of $\{N_{ij}\}$ and $\Sigma^m(0)$, and $\{\tilde{N}_{ij}\}$ and Σ^m are functions of $\{N_{ij}\}$ and $\Sigma^m(0)$, it follows that $\{U_{j0}\}$, $\{U_{ijk}\}$ and $\{X_j(0)\}$ are independent of $\{N_{ij}\}$, Σ^m and $\{\tilde{N}_{ij}\}$. Since $\{U_{j0}\}$ and $\{U_{ijk}\}$ are independent and identically distributed and $\{X_j(0)\}$ is exchangeable, by Lemmas A5.1 and A5.2 (in the Appendix), $\{\tilde{U}_{j0}\}$, $\{\tilde{U}_{ijk}\}$ and $\{Y_j^m(0)\}$ are independent of $\{\tilde{N}_{ij}\}$ and Σ^m, the $\tilde{U}_{j0}$ and the $\tilde{U}_{ijk}$ have the same distribution as U_{j0} and U_{ijk} and $(Y_1^m(0), \dots, Y_m^m(0))$ has the same distribution as $(X_1(0), \dots, X_m(0))$. Note that using this particular choice of $\{\tilde{N}_{ij}\}$, $\{\tilde{U}_{j0}\}$ and so forth in (2.3), $Y_j^m(t) = X^m_{\Sigma_j^m(t)}$ and part (b) follows.

Finally, (2.9) follows by the construction of Y^m and Σ^m. As in the proof of part (b), the σ-algebra $\mathscr{E}$ generated by $\{\tilde{U}_{j0}, 1 \le j \le m\}$, $\{\tilde{U}_{ijk}, i \le m\}$, $\{U_{j0}, j > m\}$, $\{U_{ijk}, i > m\}$, $\{Y_j^m(0), j \le m\}$, $\{X_j(0), j > m\}$ and $\{N_{ij}, m < i < j\}$ is independent of $\{\tilde{N}_{ij}, i, j \le m\}$, $\{\hat{N}_{ij}, i \le m < j\}$ and Σ^m. Since $\mathscr{H}_t^m \subset \mathscr{G}_t \vee \mathscr{E}$, $\Sigma^m(t)$ is independent of $\mathscr{H}_t^m$ by part (a). □

Note that the process $Y^m = (Y_1^m, \dots, Y_m^m)$ constructed above has the same empirical measure as $(X_1, \dots, X_m)$.

THEOREM 2.2. *For each* $t > 0$, $(X_1(t), X_2(t), \dots)$ *is exchangeable.*

PROOF. It is enough to show that for each m, $(X_1(t), \dots, X_m(t))$ is exchangeable. Since in the above construction, $\Sigma^m(t)$ is independent of $\mathscr{G}_t$ (as defined in Lemma 2.1), $\Sigma^m(t)$ must also be independent of $Y^m(t)$ and the exchangeability of $(X_1(t), \dots, X_m(t))$ follows from the exchangeability of $(Y_1^m(t), \dots, Y_m^m(t))$. □

We will need the following lemma concerning the Fleming–Viot process. Let $\mathscr{G}_t^0 = \sigma(Z(s)\colon s \le t)$.

LEMMA 2.3. *Let* Z *be a Fleming–Viot process with generator* $\mathbb{A}$, *let* $g \in \mathscr{D}(B)$ *and define* $S(t) = \langle g, Z(t)\rangle$. *Then* S *has a continuous modification and*

$$M(t) = S(t) - \int_0^t \langle Bg, Z(s)\rangle\, ds \tag{2.14}$$

is a $\{\mathscr{G}_t^0\}$ *martingale with quadratic variation*

$$\int_0^t \big(\langle g^2, Z(s)\rangle - \langle g, Z(s)\rangle^2\big)\, ds. \tag{2.15}$$

In particular, for each $f \in \overline{C}^2(\mathbb{R})$,

$$\begin{aligned} f(S(t)) - \int_0^t \Big(\tfrac{1}{2}\big(\langle g^2, Z(s)\rangle - \langle g, Z(s)\rangle^2\big) f''(S(s)) \\ + \langle Bg, Z(s)\rangle f'(S(s))\Big)\, ds \\ = \int_0^t f'(S(s))\, dM(s) \end{aligned} \tag{2.16}$$

is a martingale.

Let $\tilde{W}$ be a Brownian motion independent of Z. Then there exists a Brownian motion W such that

$$\begin{aligned}\langle g, Z(t)\rangle = \langle g, Z(0)\rangle &+ \int_0^t \sqrt{\langle g^2, Z(s)\rangle - \langle g, Z(s)\rangle^2}\, dW(s)\\ &+ \int_0^t \langle Bg, Z(s)\rangle\, ds.\end{aligned}\tag{2.17}$$

PROOF. The fact that M is a martingale follows by applying $\mathbb{A}$ to $G(\mu) = \langle g, \mu\rangle$. The existence of a continuous modification follows as in Ethier and Kurtz [(1987), Lemma 2.1]. The fact that (2.15) gives the quadratic variation follows by noting that $\mathbb{A}G^2(\mu) - 2G(\mu)\mathbb{A}G(\mu) = \langle g^2, \mu\rangle - \langle g, \mu\rangle^2$, and (2.16) follows by Itô's formula. Define $Q(t) = \langle g^2, Z(t)\rangle - \langle g, Z(t)\rangle^2$. Then

$$W(t) = \int_0^t \frac{1}{\sqrt{Q(s)}} I_{\{Q(s)>0\}}\, dM_g(s) + \int_0^t I_{\{Q(s)=0\}}\, d\tilde{W}(s)\tag{2.18}$$

is a Brownian motion satisfying (2.17). □

THEOREM 2.4. *Let $\{f_k\} \subset \mathscr{D}(B)$, with $\|f_k\| \le 1$ for all k, be separating for $\mathscr{P}(E)$ (see Lemma 1.1) and define a metric on $\mathscr{P}(E)$ by*

$$q(\mu, \nu) = \sum_{k=1}^{\infty} \frac{1}{2^k} \left| \int f_k\, d\mu - \int f_k\, d\nu \right|, \qquad \mu, \nu \in \mathscr{P}(E).\tag{2.19}$$

Let

$$Z_m(t) = \frac{1}{m} \sum_{i=1}^{m} \delta_{X_i(t)} = \frac{1}{m} \sum_{i=1}^{m} \delta_{Y_i^m(t)}\tag{2.20}$$

and define $\mathscr{H}_t^m = \sigma(Z_m(s), (X_{m+1}(s), X_{m+2}(s), \ldots), s \le t)$ and $\mathscr{H}_t = \bigcap \mathscr{H}_t^m$. Then there exists a q-continuous, $\mathscr{P}(E)$-valued process Z such that for each $t > 0$,

$$\lim_{m\to\infty} Z_m(t) = Z(t) \quad a.s.\tag{2.21}$$

in the weak topology,

$$\lim_{m\to\infty} \sup_{s\le t} q(Z_m(s), Z(s)) = 0 \quad a.s.\tag{2.22}$$

and Z is a Fleming–Viot process with generator $\mathbb{A}$.

For $1 \le k \le m$, $f \in B(E^k)$ and $y \in E^m$, define

$$\Gamma_{km} f(y) \equiv \frac{(m-k)!}{m!} \sum_{1\le i_1 \ne \cdots \ne i_k \le m} f(y_{i_1}, \ldots, y_{i_k}).\tag{2.23}$$

Then

$$E\left[f(X_1(t), \ldots, X_k(t)) | \mathscr{H}_t^m\right] = \Gamma_{km} f(Y^m(t))\tag{2.24}$$

and, with Z^k denoting the k-fold product of Z,

$$E\left[f(X_1(t), \ldots, X_k(t)) | \mathscr{H}_t\right] = \langle f, Z^k(t)\rangle.\tag{2.25}$$

In particular,

$$P\{X_k(t) \in G | \mathscr{H}_t^m\} = Z_m(t, G)\tag{2.26}$$

and

$$P\{X_k(t) \in G | \mathscr{H}_t\} = Z(t, G). \tag{2.27}$$

REMARK 2.5. Typically, convergence in q will imply weak convergence (in particular if the mutation process is a Feller process), but not necessarily. For example, let $E = \mathbb{R}$ and suppose the mutation process decreases at rate 1 until it hits an integer, say k, and then jumps to a point uniformly distributed on the interval $(k, k+1]$. Then $Bf(x) = -f'(x)$ for f with bounded left derivatives satisfying $f(k+) = \int_k^{k+1} f(x)\,dx$, for each integer k. Note, in particular, that the Fleming–Viot process with this mutation operator does not have sample paths continuous in the weak topology.

PROOF OF THEOREM 2.4. By Lemma 2.1, Σ^m is independent of $\mathscr{H}_t^m$ which implies (2.24). Since $\mathscr{H}_t^m \supset \mathscr{H}_t^{m+1}$, for fixed t and G, the left side of (2.24) is a reverse martingale and converges a.s. to (2.25). In particular,

$$E[f(X_k(t)) | \mathscr{H}_t^m] = \int f(x) Z_m(t, dx) \tag{2.28}$$

converges almost surely for each $f \in \overline{C}(E)$. Since E is separable, there is a countable convergence determining collection in $\overline{C}(E)$, and it follows that $Z_m(t)$ converges almost surely to $Z(t)$ in $\mathscr{P}(E)$, and (2.25) follows from (2.24). In fact, since (2.28) is a reverse martingale with uniformly bounded differences, for $\varepsilon > 0$, Lemma A3.1 gives

$$P\left\{\left|\int f(x) Z_m(t, dx) - \int f(x) Z(t, dx)\right| \geq \varepsilon\right\} \leq Ce^{-\eta m}, \tag{2.29}$$

where C and $\eta > 0$ depend only on ε and $\|f\|$. [The argument here is the standard martingale proof of de Finetti's theorem. See, for example, Durrett (1991) or Chow and Teicher (1988).]

To see that Z is a solution of the martingale problem for $\mathbb{A}$, first verify that for $f \in \mathscr{D}(B) \cap B(E^k)$, $A_m \Gamma_{km} f = \Gamma_{km} Af$. From (2.24), (2.25) and (1.16) it follows that

$$\begin{aligned}
0 &= E\bigg[E\bigg[\Gamma_{km} f(Y_m(t+s)) - \Gamma_{km} f(Y_m(t)) \\
&\qquad\qquad - \int_t^{t+s} A_m \Gamma_{km}(Y_m(u))\,du \bigg| \mathscr{H}_t^m\bigg]\bigg| \mathscr{H}_t\bigg] \\
&= E\bigg[f(X_1(t+s), \ldots, X_k(t+s)) - f(X_1(t), \ldots, X_k(t)) \\
&\qquad\qquad - \int_t^{t+s} Af(X_1(u), \ldots, X_k(u))\,du \bigg| \mathscr{H}_t\bigg] \\
&= E\bigg[\langle f, Z(t+s)^k\rangle - \langle f, Z(t)^k\rangle - \int_t^{t+s} \langle Af, Z(u)^k\rangle\,du \bigg| \mathscr{H}_t\bigg] \\
&= E\bigg[F(Z(t+s)) - F(Z(t)) - \int_t^{t+s} \mathbb{A}F(Z(u))\,du \bigg| \mathscr{H}_t\bigg]
\end{aligned} \tag{2.30}$$

and it follows that Z is a solution of the martingale problem for $\mathbb{A}$.

Lemma 2.3 ensures that for $f \in \mathscr{D}(B)$, $\int f(x) Z(t, dx)$ has a continuous modification, so in particular, Z has a q-continuous modification.

Let $R_i(t,h) = 1$ if X_i "looks down" during the time interval $(t, t+h]$ and $R_i(t,h) = 0$ otherwise. Note $P(R_i(t,h) = 0\} = e^{-(i-1)h}$. For $f \in \mathscr{D}(B)$ and $\varepsilon > 0$,

$$
\begin{aligned}
&P\left\{ \sup_{t \le s \le t+h} \left| \int f(x) Z_m(s, dx) - \int f(x) Z_m(t, dx) \right| \ge \varepsilon \right\} \\
&\quad = P\left\{ \sup_{t \le s \le t+h} \left| \frac{1}{m} \sum_{i=1}^m f(X_i(s)) - \frac{1}{m} \sum_{i=1}^m f(X_i(t)) \right| \ge \varepsilon \right\} \\
&\quad \le P\left\{ \sup_{t \le s \le t+h} \left| \frac{1}{m} \sum_{i=1}^m \left(f(X_i(s)) - f(X_i(t)) \right.\right.\right. \\
&\qquad\qquad \left.\left.\left. - \int_t^s Bf(X_i(u))\, du \right) (1 - R_i(t,h)) \right| > \frac{\varepsilon}{4} \right\} \qquad (2.31) \\
&\qquad + P\left\{ (2\|f\| + h\|Bf\|) \left| \frac{1}{m} \sum_{i=1}^m \left(R_i(t,h) - 1 + e^{-(i-1)h} \right) \right| > \frac{\varepsilon}{4} \right\} \\
&\qquad + P\left\{ \frac{1}{m} \sum_{i=1}^m \int_t^{t+h} |Bf(X_i(u))|\, du > \frac{\varepsilon}{4} \right\} \\
&\qquad + P\left\{ (2\|f\| + h\|Bf\|) \frac{1}{m} \sum_{i=1}^m (1 - e^{-(i-1)h}) \ge \frac{\varepsilon}{4} \right\}.
\end{aligned}
$$

The independence of the R_i from the evolution of the X_i between look-downs implies that the process in the first term on the right is a martingale. Defining $\hat{\varphi}(x, \lambda) = e^{-\lambda x} + e^{\lambda x}$, Doob's inequality bounds this term by

$$
\begin{aligned}
&\inf_{\lambda > 0} \left\{ E\left[\hat{\varphi}\left(\frac{1}{m} \sum_{i=1}^m \left(f(X_i(t+h)) - f(X_i(t)) \right.\right.\right.\right. \\
&\qquad \left.\left.\left. - \int_t^{t+h} Bf(X_i(u))\, du \right) (1 - R_i(t,h), \lambda) \right] \qquad (2.32) \\
&\qquad\qquad \left. \times [\hat{\varphi}(\varepsilon, \lambda)]^{-1} \right\}.
\end{aligned}
$$

It follows as in the proof of Theorem 5.9 of Kurtz (1972) that there exist C and $\eta > 0$ depending only on ε and $2\|f\| + h\|Bf\|$ (a constant bounding the summands) such that the first term is bounded by $Ce^{-\eta m}$. The second term is bounded by a similar expression, and the third and fourth terms are zero if $h\|Bf\| < \varepsilon/4$ and $(2\|f\| + h\|Bf\|)(1 - e^{-h(m-1)}) < \varepsilon/4$. Consequently, C and $\eta > 0$ may be selected depending only on ε, $\|f\|$ and $\|Bf\|$ such that for h sufficiently small, (2.29) holds and the left side of (2.31) is also bounded by $Ce^{-\eta m}$. Let $h_m \to 0$ slowly enough so that $\sum e^{-\eta m}/h_m < \infty$ for every $\eta > 0$ and fast enough so that $m h_m \to 0$ (e.g., $h_m = m^{-2}$). For $T > 0$, let $H_{m,T} = \{k h_m$:

$k \leq T/h_m\}$. Let $f \in \mathscr{D}(B)$, assume $\int f(x)Z(\cdot, dx)$ is continuous and define

$$D_m = \left\{ \sup_{t \leq T} \sup_{t \leq s \leq t+h_m} \left| \int f(x)Z(s,dx) - \int f(x)Z(t,dx) \right| < \varepsilon \right\}. \tag{2.33}$$

Note that $D_m \subset D_{m+1}$ and by the continuity of $\int f(x)Z(\cdot, dx)$, $P(D_m) \to 1$. For m sufficiently large (i.e., h_m sufficiently small),

$$\begin{aligned} P\left(\left\{ \sup_{t \leq T} \left| \int f(x)Z_m(t,dx) - \int f(x)Z(t,dx) \right| \geq 3\varepsilon \right\} \cap D_m \right) \\ \leq \sum_{t \in H_{m,T}} P\left\{ \left| \int f(x)Z_m(t,dx) - \int f(x)Z(t,dx) \right| \geq \varepsilon \right\} \\ + \sum_{t \in H_{m,T}} P\left\{ \sup_{t \leq s \leq t+h_m} \left| \int f(x)Z_m(s,dx) - \int f(x)Z_m(t,dx) \right| \geq \varepsilon \right\} \\ \leq \frac{2CT}{h_m} e^{-\eta m}. \end{aligned} \tag{2.34}$$

Summing over m, the right side converges, and Borel–Cantelli and the properties of D_m ensure that

$$\lim_{m \to \infty} \sup_{t \leq T} \left| \int f(x)Z_m(t,dx) - \int f(x)Z(t,dx) \right| = 0, \tag{2.35}$$

which completes the proof of the theorem. □

Corollary 2.6. *Let τ be a finite $\{\mathscr{H}_t\}$ stopping time. Then $(X_1(\tau), X_2(\tau), \dots)$ is exchangeable.*

Proof. If τ is discrete, then (2.25) implies the desired exchangeability. For the general case, let $\{\tau_n\}$ be a decreasing sequence of discrete $\{\mathscr{H}_t\}$ stopping times converging to τ. By the right continuity of the X_k, $(X_1(\tau_n), X_2(\tau_n), \dots) \to (X_1(\tau), X_2(\tau), \dots)$ and the exchangeability follows. □

We now consider a construction similar to that of Y^m above, but involving $(X_{m+1}, \dots, X_{m+n})$. Define an S_n-valued process $\tilde{\Sigma}^n$ and counting processes $\{\tilde{N}_{ij}, 1 \leq i \neq j \leq n\}$ as follows. Let $\tilde{\Sigma}^n(0)$ be uniformly distributed on S_n and independent of all other processes. Let

$$\tilde{N}_{ij}(t) = \sum_{1 \leq k < l \leq n} \int_0^t I_{\{\tilde{\Sigma}_i^n(s-)=k,\, \tilde{\Sigma}_j^n(s-)=l\}}\, dN_{m+k,m+l}(s) \tag{2.36}$$

and let $\tilde{\Sigma}^n$ be constant except for discontinuities determined by $\tilde{\Sigma}^n(\tilde{\tau}_{ijk}) = M_{ijk}\tilde{\Sigma}^n(\tilde{\tau}_{ijk}-)$, where $\tilde{\tau}_{ijk}$ is the kth jump time of $\tilde{N}_{ij}$ or, more precisely, interpreting $\tilde{\Sigma}^n$ as a $\mathbb{Z}^n$-valued process,

$$\tilde{\Sigma}^n(t) = \sum_{1 \leq i < j \leq n} \int_0^t \left(M_{i,j,\tilde{N}_{ij}(s-)+1} \tilde{\Sigma}^n(s-) - \tilde{\Sigma}^n(s-) \right) d\tilde{N}_{ij}(s). \tag{2.37}$$

Equations (2.36) and (2.37) determine the $\tilde{N}_{ij}$ and $\tilde{\Sigma}^n$ recursively. In addition define $\{\hat{N}_{ij}, 1 \le i \le m, 1 \le j \le n\}$ by

$$\hat{N}_{ij}(t) = \sum_{k=1}^{n} \int_0^t I_{\{\tilde{\Sigma}_j^n(s-)=k\}} \, dN_{i,m+k}(s). \tag{2.38}$$

As in Lemma 2.1, $\tilde{\Sigma}^n(t)$ is independent of

$$\tilde{\mathscr{G}}_t = \sigma\Big\{\tilde{N}_{ij}(s), \hat{N}_{kl}(s): s \le t, 1 \le i, j \le n, 1 \le k \le m, l \le l \le n\Big\}. \tag{2.39}$$

Define $\tilde{Y}_j^n(t) = X_{m+\tilde{\Sigma}_j^n(t)}(t)$. Then $\tilde{N}_{ij}$ counts the number of times that $\tilde{Y}_j^n$ "copies" $\tilde{Y}_i^n$, and $\hat{N}_{ij}$ counts the number of times $\tilde{Y}_j^n$ copies X_i. Setting $X^m = (X_1, \dots, X_m)$, it follows that $(X^m, \tilde{Y}^n)$ is a solution of the martingale problem for

$$\begin{aligned} A^{mn} f(x,y) = Bf(x,y) &+ \sum_{1 \le i < j \le m} \Big(f\big(\theta_{ij}(x), y\big) - f(x,y) \Big) \\ &+ \sum_{\substack{1 \le i \le m \\ m+1 \le j \le m+n}} \Big(f\big(x, \theta_j(y|x_i)\big) - f(x,y) \Big) \\ &+ \tfrac{1}{2} \sum_{m+1 \le i, j \le m+n} \Big(f\big(x, \theta_{ij}(y)\big) - f(x,y) \Big) \end{aligned} \tag{2.40}$$

for $f \in \mathscr{D}(B) \cap B(E^{m+n})$. For $g \in \mathscr{D}(B) \cap B(E^{m+k})$, define

$$\Gamma^n g(x,y) \equiv \frac{1}{n^k} \sum_{1 \le i_1 \ne \cdots \ne i_k \le n} g(x, y_{i_1}, \dots, y_{i_k}). \tag{2.41}$$

Applying A^{mn} we obtain

$$\begin{aligned} A^{mn} \Gamma^n g = \Gamma^n Bg &+ \sum_{1 \le i < j \le m} \Big(\Gamma^n g\big(\theta_{ij}(\cdot), \cdot\big) - \Gamma^n g \Big) \\ &+ \sum_{\substack{1 \le i \le m \\ m+1 \le j \le m+k}} \Big(\Gamma^n g\big(\cdot, \theta_j(\cdot|x_i)\big) - \Gamma^n g \Big) \\ &+ \tfrac{1}{2} \sum_{m+1 \le i \ne j \le m+k} \Gamma^n(\Phi_{ij} g - g). \end{aligned} \tag{2.42}$$

Noting that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{\tilde{Y}_i^n(t)} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{X_{m+i}(t)} = Z(t) \quad \text{a.s.} \tag{2.43}$$

and in particular that

$$\lim_{n\to\infty} \Gamma^n g(X^m(t), Y^m(t)) = \langle g(X^m(t), \cdot), Z(t)^k \rangle \quad \text{a.s.,}$$

we have the following theorem.

THEOREM 2.7. *Let X and Z be as in Theorem* 2.4. *For $g \in \mathscr{D}(B) \cap B(E^{m+k})$ and $(x, \mu) \in E^m \times \mathscr{P}(E)$, define $G(x, \mu) = \langle g(x, \cdot), \mu^k \rangle$ and*

$$\begin{aligned}\mathbb{A}^m G(x, \mu) = \left\langle Bg(x, \cdot), \mu^k \right\rangle &+ \sum_{1 \le i < j \le m} \left\langle g\left(\theta_{ij}(x), \cdot\right) - g(x, \cdot), \mu^k \right\rangle \\ &+ \sum_{\substack{1 \le i \le m \\ m+1 \le j \le m+k}} \left\langle g\left(x, \theta_j(\cdot|x_i)\right) - g(x, \cdot), \mu^k \right\rangle \\ &+ \sum_{m+1 \le i < j \le m+k} \langle \Phi_{ij} g(x, \cdot) - g(x, \cdot), \mu^k \rangle.\end{aligned} \tag{2.44}$$

Then $(X_1, \ldots, X_m, Z)$ is a solution of the martingale problem for $\mathbb{A}^m$.

The martingale problem for $\mathbb{A}^m$ is well posed (*i.e., existence and uniqueness hold for each initial distribution*) *and it follows that $(X_1, \ldots, X_m, Z)$ is a Markov process.*

PROOF. For $g \in \mathscr{D}(B) \cap B(E^{m+k})$,

$$\Gamma^n g\left(X^m(t), \tilde{Y}^n(t)\right) - \int_0^t A^{mn} \Gamma^n g\left(X^m(s), \tilde{Y}^n(s)\right) ds \tag{2.45}$$

is a martingale with respect to the filtration $\mathscr{G}_t^{mn} = \sigma(X^m(s), \tilde{Y}^m(s): s \le t)$. Each term in (2.45) converges a.s. to the corresponding term in

$$G(X^m(t), Z(t)) - \int_0^t \mathbb{A}^m G(X^m(s), Z(s))\, ds. \tag{2.46}$$

[The elimination of the $\frac{1}{2}$ and the change in the range of the summation is permitted by the fact that $\langle \Phi_{ij} g(x, \cdot) - g(x, \cdot), \mu^k \rangle = \langle \Phi_{ji} g(x, \cdot) - g(x, \cdot), \mu^k \rangle$.] The assertion that $(X_1, \ldots, X_m, Z)$ is a solution of the martingale problem for $\mathbb{A}^m$ then follows by Theorem 4.8.10 of Ethier and Kurtz (1986).

Observe that for $g \in \mathscr{D}(B) \cap B(E^{m+k})$, $\mathbb{A}^m G(x, \mu) = \langle Ag(x, \cdot), \mu^k \rangle$. Uniqueness of solutions of the martingale problem then follows from Corollary 3.3 of Kurtz (1995), and the Markov property follows from uniqueness by Theorem 4.4.2 of Ethier and Kurtz (1986). □

Theorem 2.7 gives us the following generalization of Lemma 2.3. Let $\mathscr{G}_t^m = \sigma(X^m(s), Z(s): s \le t)$.

COROLLARY 2.8. *Let Z be a Fleming–Viot process with generator $\mathbb{A}$, let $g \in \mathscr{D}(B)$ and define $S(t) = \langle g, Z(t) \rangle$. Then for each $m > 0$,*

$$M(t) = S(t) - \int_0^t \left(\langle Bg, Z(s) \rangle + \sum_{i=1}^m \left(g(X_i(s)) - S(s) \right) \right) ds \tag{2.47}$$

is a $\{\mathscr{G}_t^m\}$ martingale with quadratic variation

$$\int_0^t \left(\langle g^2, Z(s) \rangle - \langle g, Z(s) \rangle^2 \right) ds. \tag{2.48}$$

Let $\tilde{W}$ be a Brownian motion independent of X^m and Z. Then there exists a Brownian motion W such that

$$\langle g, Z(t)\rangle = \langle g, Z(0)\rangle + \int_0^t \sqrt{\langle g^2, Z(s)\rangle - \langle g, Z(s)\rangle^2}\, dW(s)$$

$$+\int_0^t \left(\langle Bg, Z(s)\rangle + \sum_{i=1}^m \left(g(X_i(s)) - \langle g, Z(s)\rangle\right)\right) ds. \tag{2.49}$$

PROOF. The fact that M is a $\{\mathscr{G}_t^m\}$ martingale follows immediately from Theorem 2.7. The quadratic variation does not depend on the filtration, so (2.48) follows from Lemma 2.3. The proof of (2.49) is the same as the proof of (2.17). □

Stationary versions. If B has a stationary distribution, then the construction of the process X can be carried out on the doubly infinite time interval. In this construction, $\{X_1(t), -\infty < t < \infty\}$ is a stationary Markov process with generator B and the N_{ij} are Poisson point processes on $(-\infty, \infty)$ with intensity 1. Associating each point in N_{ij} with an independent version of U, the construction is essentially the same as before. Note that the stationarity of X_1 and the increments of the N_{ij} imply the stationarity of X and Z. In particular, we see that the existence of a stationary version X_1 of the mutation process implies the existence of a stationary version of the corresponding Fleming–Viot process. Furthermore, if X_1 is ergodic, then Z is ergodic. See Section 4.

THEOREM 2.9. *Suppose that the mutation process is an $\mathbb{R}^d$-valued process with stationary independent increments (e.g., Brownian motion or a continuous time random walk). Let X be as above, and define X^n for $t \geq -n$ by $X_k^n(t) = X_k(t+n) - X_1(n)$. Then X^n converges in distribution to a process X^∞ defined on the doubly infinite time interval, and the $(\mathbb{R}^d)^\infty$-valued process $\hat{X} = (0, X_2^\infty - X_1^\infty, X_3^\infty - X_1^\infty, \dots)$ is stationary and ergodic.*

REMARK 2.10. (a) A similar limit theorem can be proved for $\tilde{X}_k^n(t) = X_k(t+n) - \int xZ(t+n, dx)$. See Dawson and Hochberg (1982) and Shiga (1982) for analogous results. Similar ideas of centering first appeared in Moran (1975, 1976) and Kingman (1976).

(b) Note that X^∞ can be constructed directly in the same way as the stationary particle model constructed above starting with a process with stationary independent increments X_1^∞ defined on $(-\infty, \infty)$ satisfying $X_1^\infty(0) = 0$.

PROOF OF THEOREM 2.9. Convergence in distribution of X^n follows from convergence in distribution of the first component $X_1^n(t) = X_1(t+n) - X_1(n)$ which is immediate, since the joint distribution of $X_1^n(t_1), \dots, X_1^n(t_m)$ does not depend on n for $n \geq \max_i(-t_i)$. □

3. Genealogical structure. Assume that the demographic process (by which we mean the process of "look-downs" determined by the N_{ij}) is defined on the time interval $(-\infty, \infty)$. For $s < t$, the ancestor at time s of the particle at level j at time t will be the particle at the level i determined by following the look-down processes backward in time. To be precise, define $N_j(a, b] = \sum_{i<j} N_{ij}(a, b]$ (where $N_{ij}(a, b]$ denotes the number of points in N_{ij} falling in the time interval $(a, b]$). Let $\gamma_j(t)$ be the time of the most recent look-down from the jth level, that is, $\gamma_j(t) = \sup\{u: N_j(u, t] > 0\}$, and let $\alpha_j(\gamma_j(t))$ be the level i such that $\gamma_j(t) \in N_{ij}$. Define $a_j(s, t) = j$ for $\gamma_j(t) \le s < t$ and $a_j(s, t) = \alpha_j(\gamma_j(t))$ for $\gamma_{\alpha_j(\gamma_j(t))}(\gamma_j(t)) \le s < \gamma_j(t)$, and extend the definition $a_j(s, t)$ to all $s < t$ in the obvious manner. Then $a_j(s, t)$ gives the level of the ancestor at time s of the particle at level j at time t.

Let $\Gamma_n(s, t) = \{a_j(s, t): j = 1, \dots, n\}$; that is, $\Gamma_n(s, t)$ is the collection of indices of particles at time s that have descendants among the first n particles at time t. Let $|\Gamma_n(s, t)|$ denote the cardinality of $\Gamma_n(s, t)$. For an arbitrary but fixed $\tilde{t}$ and $u \ge 0$, define $D_n(u) = |\Gamma_n(\tilde{t} - u, \tilde{t})|$ and define an equivalence relation $R^n(u)$ on $\{1, \dots, n\}$ by $(i, j) \in R^n(u)$ if and only if $a_i(\tilde{t} - u, \tilde{t}) = a_j(\tilde{t} - u, \tilde{t})$. Thus the equivalence classes of $R^n(u)$ consist of the levels of particles at time $\tilde{t}$ that have the same ancestor at time $\tilde{t} - u$.

Recall that an n-coalescent is a continuous time Markov chain with state space $\mathscr{S}_n$, the set of equivalence relations on $\{1, \dots, n\}$, initial value $\{(i, i): i = 1, \dots, n\}$ and transition intensities, for $\xi \ne \eta \in \mathscr{S}_n$,

$$q_{\xi\eta} = \begin{cases} 1, & \text{if } \eta \text{ is obtained by coalescing two of the equivalence} \\ & \quad \text{classes of } \xi, \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

Theorem 3.1. *The process $\{R^n(u), u \ge 0\}$ is an n-coalescent and $D_n(u) = |\Gamma_n(\tilde{t} - u, \tilde{t})| = |R^n(u)|$, the number of equivalence classes in $R^n(u)$, is a pure death Markov chain with transition intensities*

$$q_{k,k-1} = \frac{k(k-1)}{2}, \qquad k = 2, \dots, n. \tag{3.2}$$

Proof. Since $a_i(\tilde{t}, \tilde{t}) = i$, $i = 1, \dots, n$, $R^n(0) = \{(i, i): i = 1, \dots, n\}$. The above construction naturally associates a level with each equivalence class of $R^n(u)$, the common value of $a_i(\tilde{t} - u, \tilde{t})$ for each value of i in the equivalence class, that is, the level of the ancestor at time $\tilde{t} - u$ of the equivalence class. Transitions occur in $R^n(u)$ exactly when the ancestor of one equivalence class looks down to the level associated with one of the other classes. For each pair of classes an event of this sort happens at rate 1. The effect of such an event is to coalesce the two equivalence classes.

In particular, for small h, $D_n(u + h) = D_n(u) - 1$ if there is a look-down from one of the levels in $\Gamma_n(\tilde{t} - u, \tilde{t})$ to one of the other levels in $\Gamma_n(\tilde{t} - u, \tilde{t})$ in the time interval $(\tilde{t} - u - h, \tilde{t} - u]$. The probability of a look-down occurring in this interval for a particular pair is $h + o(h)$, so if $D_n(u) = k$, the probability of a look-down occurring for some pair in $\Gamma_n(\tilde{t} - u, \tilde{t})$ is just $\binom{k}{2}h + o(h)$. The Markov property for R^n and D_n is an immediate consequence of the independence of the increments of the Poisson processes used in our construction. □

Let $\tau_k^n = \inf\{u: D_n(u) = k\}$. Then

$$E[\tau_k^n] = \sum_{m=k+1}^{n} \frac{2}{m(m-1)} = \frac{2}{k} - \frac{2}{n} \tag{3.3}$$

which converges as $n \to \infty$. It follows that $D(u) = \lim_{n\to\infty} D_n(u) < \infty$ for all $u > 0$, and hence that $\Gamma(s,t)$, the collection of indices of particles at time s that have a descendent at some level at time t, is a finite set. In particular, the equivalence relation $R(u)$, $u \geq 0$, on the natural numbers $\mathbb{N}$ defined by $(i, j) \in R(u)$ if and only if $a_i(\tilde{t} - u, \tilde{t}) = a_j(\tilde{t} - u, \tilde{t})$ has the property that for all $u > 0$ it determines a finite collection of equivalence classes.

THEOREM 3.2. *The process* $\{R(u), u \geq 0\}$ *is a coalescent.*

PROOF. This assertion is immediate from the previous theorem, since the coalescent on $\mathbb{N}$ is defined to be a process of equivalence relations on $\mathbb{N}$ with the property that for each n, its restriction to $\{1, \dots, n\}$ gives an n-coalescent. □

The two major new tools in population genetics of the last decade have been the uses of genealogy and measure-valued diffusions. There are a variety of limit theorems which show that appropriate genealogical or measure-valued processes arise as robust descriptions of certain classes of discrete models. One consequence of our construction is that these two previously somewhat disparate approaches are unified by the embedding of the Fleming–Viot process and the familiar genealogical processes in the particle model. In particular,the genealogical tree with mutation considered by Donnelly and Tavaré (1987a) can be embedded in the particle model. Another consequence is an explicit construction of the coalescent. Kingman (1982b) uses "paintboxes" to construct its jump chain.

We collect various properties of R in the following proposition. All are properties of the coalescent [Kingman (1982b)] and so follow from the previous theorem.

PROPOSITION 3.3. (a) *The process* D *is a Markov death process on* $\mathbb{N}$ *with an entrance boundary at* ∞ *and death rate* $\binom{k}{2}$ *from* k, $D(0) = \infty$, $D(u) < \infty$ *a.s. for* $u > 0$ *and* $\lim_{u\to\infty} D(u) = 1$ *a.s. For* $\tau_k \equiv \inf\{u: D(u) = k\}$, $E[\tau_k] = 2/k$.

(b) D *is independent of the jump chain of* R.

(c) *For each* u, $R(u)$ *is exchangeable* (*that is*, *its distribution is invariant under relabelings of the levels at time* $\tilde{t}$), *as is* $R_k \equiv R(\tau_k)$, *for each* k. *It* *follows that for $u > 0$ (or k finite) all the equivalence classes of $R(u)$ (or R_k) are infinite and in fact each class has positive density a.s., that is, if C is such an equivalence class,*

$$\chi_C = \lim_{n \to \infty} n^{-1} |C \cap \{1, \ldots, n\}| \tag{3.4}$$

exists and is positive a.s.

(d) *Let $C_1(u), \ldots, C_{D(u)}(u)$ denote the equivalence classes of $R(u)$, labeled in a way that does not depend on their composition. Then for $j_1 \neq j_2 \neq \cdots \neq j_m$,*

$$P\left\{j_k \in C_{i_k}(u), k = 1, \ldots, m | D(u), \chi_{C_1(u)}, \ldots, \chi_{C_{D(u)}(u)}\right\} = \prod_{k=1}^{m} \chi_{C_{i_k}(u)}. \tag{3.5}$$

3.4. REMARK. The statement in (d), due to Kingman (1982a), is an analog of de Finetti's theorem for exchangeable random equivalence relations. Essentially, it says that the interesting randomness in $R(u)$ is contained in the number of equivalence classes and their sizes.

In the sequel it will be useful to consider the coalescent with time running in the opposite direction, that is, to reverse time about the point $\tilde{t} - \tau_1$ at which all of the particles at time $\tilde{t}$ first have a common ancestor. For definiteness, take $\tilde{t} = 0$. Define $\tilde{R}^\infty(s) = R((\tau_1 - s) -)$, where $R(u -)$ denotes the left limit at u. We use the left limit in order to make $\tilde{R}^\infty$ right continuous. The usual direction of time for the coalescent is with time running backward in our original (particle) construction, so we now consider time running forward from $-\tau_1$. Transitions now involve the splitting of equivalence classes. In view of Proposition 3.3(b), we may consider separately the time reversal of the process D, which determines the time points at which the coalescent jumps, and that of its jump chain $\{R_k\}$. The former, $\tilde{D}^\infty(s) = D((\tau_1 - s) -)$, is easy. It is simply a Markov birth process with $\tilde{D}^\infty(0) = 2$ and birth rates $\binom{k}{2}$ from state k. For finite n, we consider $\tilde{R}^n(s) = R^n((\tau_1^n - s) -)$, the reversal of the n-coalescent around the time at which the first n particles at time 0 first have a common ancestor. Analogous observations hold for $\tilde{R}^\infty$, the time reversal of R^∞.

From Kingman (1982b), the sequence $R_1, R_2, \ldots$ is Markov with the structure $R_1 = \{(i, j): i, j = 1, 2, \ldots\}$, and the transition from R_k to R_{k+1} involves the splitting of an equivalence class of R_k. The class to be split, C, is chosen with probability equal to its size, χ_C. It is split into two classes C' and $C - C'$ of sizes $U\chi_C$ and $(1 - U)\chi_C$, where U is uniformly distributed on $[0, 1]$ independent of everything else, and conditional on U, each element of C is independently assigned to C' with probability U.

If $\{R_k^n\}$ denotes the jump chain of the n-coalescent R^n, then again the sequence is Markov with $R_1^n = \{(i, j): i, j = 1, \ldots, n\}$ and transition probabilities

$$\begin{aligned} &P\{R_{k+1}^n = (\xi_1, \ldots, \xi_{i-1}, \xi^*, \xi_i - \xi^*, \xi_{i+1}, \ldots, \xi_k) | R_k^n = (\xi_1, \ldots, \xi_k)\} \\ &\quad = \frac{2}{n - k} \binom{|\xi_i|}{|\xi^*|}^{-1}, \qquad \xi^* \subset \xi_i, \xi^* \neq \xi_i, \end{aligned} \tag{3.6}$$

where $R_k^n = (\xi_1, \dots, \xi_k)$ means that the equivalence classes of R_k^n are $\xi_1, \dots, \xi_k$. The dynamics of the sequence $\{R_k^n\}$ are the discrete analog of those of $\{R_k\}$. In the transition from R_k^n to R_{k+1}^n, a particular class ξ_i will be split with probability $(|\xi_i| - 1)/(n-k)$. The splitting consists of partitioning its size uniformly (i.e., choosing an ordered pair uniformly from $\{(a, b)\colon a + b = |\xi_i|\}$) and then choosing uniformly one of the possible assignments of the individuals in ξ_i to two classes of sizes a and b.

Since the transition probabilities for $\{R_k^n$ (or $\{R_k\}$) depend only on the sizes of the equivalence classes, it follows that the process $(\tilde{D}^n, S_{1_2}^n, \dots, S_{\tilde{D}^n}^n)$, where the S_i^n are the cardinalities of the equivalence classes in $\tilde{R}^n$, is Markovian and has generator

$$\begin{aligned} &G^n f(k, s_1, \dots, s_k) \\ (3.7) \qquad &= \binom{k}{2} \sum_{i=1}^{k} \frac{1}{(n-k)} \sum_{m=1}^{s_i - 1} \bigl(f(k+1, s_1, \dots, s_{i-1}, m, \\ &\qquad\qquad s_i - m, s_{i+1}, \dots, s_k) - f(k, s_1, \dots, s_k) \bigr), \end{aligned}$$

and the process $(\tilde{D}^\infty, \chi_1, \dots, \chi_{\tilde{R}^\infty})$, where the χ_i are the sizes of the equivalence classes as defined in (3.4), is Markovian with generator

$$\begin{aligned} &G^\infty f(k, \chi_1, \dots, \chi_k) \\ (3.8) \qquad &= \binom{k}{2} \sum_{i=1}^{\infty} \chi_i \int_0^1 \bigl(f(k+1, \chi_1, \dots, \chi_{i-1}, u\chi_i, (1-u)\chi_i, \\ &\qquad\qquad \chi_{i+1}, \dots, \chi_k) - f(k, \chi_1, \dots, \chi_k) \bigr)\, du. \end{aligned}$$

[Observe that G^∞ can be obtained as the limit of the G^n under the map $(k, s_1, s_k) \to (k, s_1/n, \dots, s_k/n)$.]

Now assume that the particle model is defined for all time $(-\infty, \infty)$ as, for example, in Theorem 2.9, where the mutation process is either stationary or a process with independent increments. Associate with each equivalence class C_i of $\tilde{R}^n(s)$ [or $\tilde{R}(s)$], the type $\tilde{X}_i(s)$ of the common ancestor for that equivalence class. That is, $\tilde{X}_i(s)$ is the type of $X_{a_j(s - \tau_1^n, 0)}(s - \tau_1^n)$ for $j \in C_i$. (Note that C_i and $\tilde{X}_i$ depend on n; however, the meaning should be clear from context and we suppress the superscript.) Since the mutation process between look-downs is independent of the demography, it follows easily that for finite n, $(\tilde{D}^n, S_1^n, \dots, S_{\tilde{D}^n}^n, \tilde{X}_1, \dots, \tilde{X}_{\tilde{D}^n})$, as a process in $\mathscr{S}^n = \{(k, \underset{\sim}{s}, \underset{\sim}{x})\colon s \in \mathbb{N}^k,\ \sum s_i = n,\ x \in E^k,\ k = 2, \dots, n\}$, is, up to time τ_1^n, Markovian with generator

$$\begin{aligned} &G^n f(k, s, x) \\ &\quad = \sum_{i=1}^{k} B_i f(k, \underset{\sim}{s}, \underset{\sim}{x}) + \binom{k}{2} \sum_{i=1}^{k} \frac{1}{(n-k)} \\ (3.9) \qquad &\quad \times \sum_{m=1}^{s_i - 1} \bigl(f(k+1, s_1, \dots, s_{i-1}, m, s_i - m, s_{i+1}, \dots, s_k, \\ &\qquad\qquad x_1, \dots, x_{i-1}, x_i, x_i, x_{i+1}, \dots, x_k) - f(k, \underset{\sim}{s}, \underset{\sim}{x}) \bigr), \end{aligned}$$

where B_i denotes the mutation operator applied to f as a function of x_i. [Define $G^n f(n, \underset{\sim}{s}, \underset{\sim}{x}) = 0$.] Note that when the ith equivalence class splits, the two new equivalence classes have type x_i, but after the split the types evolve independently. The analogous generator for $(\tilde{D}^\infty, \chi_1, \dots, \chi_{\tilde{R}^\infty}, \tilde{X}_1, \dots, \tilde{X}_{\tilde{R}^\infty})$ in $\mathscr{S}^\infty = \{(k, \underset{\sim}{\chi}, \underset{\sim}{x})\colon \chi_i \geq 0, \sum \chi_i = 1, \underset{\sim}{x} \in E^k, k = 2, 3, \dots, \infty\}$ is

$$\begin{aligned} G^\infty f(k, \underset{\sim}{\chi}, \underset{\sim}{x}) = {} & \sum_{i=1}^{k} B_i f(k, \underset{\sim}{\chi}, \underset{\sim}{x}) \\ & + \binom{k}{2} \sum_{i=1}^{k} \chi_i \int_0^1 \bigl(f(k+1, \chi_1, \dots, \chi_{i-1}, u\chi_i, (1-u)\chi_i, \\ & \qquad \chi_{i+1}, \dots, \chi_k, x_1, \dots, x_{i-1}, x_i, x_i, \\ & \qquad x_{i+1}, \dots, x_k) - f(k, \underset{\sim}{\chi}, \underset{\sim}{x}) \bigr)\, du \end{aligned} \tag{3.10}$$

for $k < \infty$ and $G^\infty f(\infty, \underset{\sim}{\chi}, \underset{\sim}{x}) = 0$.

Let $\hat{\mathscr{S}}^n = \{(k, \underset{\sim}{x})\colon \underset{\sim}{x} \in E^k, k = 2, 3, \dots, n\}$ and define $\hat{G}^n$ by

$$\begin{aligned} \hat{G}^n f(k, \underset{\sim}{x}) = {} & \sum_{i=1}^{k} B_i f(k, \underset{\sim}{x}) + I_{[2,n)}(k) \frac{k-1}{2} \\ & \times \sum_{i=1}^{k} \bigl(f(k+1, x_1, \dots, x_{i-1}, x_i, x_i, x_{i+1}, \dots, x_k) \\ & \qquad - f(k, \chi, \underset{\sim}{x}) \bigr) \end{aligned} \tag{3.11}$$

for $k < \infty$ and $\hat{G} f(\infty, \underset{\sim}{x}) = 0$.

Let $\Pi_{k,n} = \{(s_1, \dots, s_k)\colon 1 \leq s_i \leq n - k, \sum s_i = n\}$. For $s \in \Pi_{k,n}$, define $\pi_{k,n}(s) = \binom{n-1}{k-1}^{-1}$, that is, the uniform distribution on $\Pi_{k,n}$.

THEOREM 3.5. *Assume, as above, that the particle model is defined for all time* $(-\infty, \infty)$. *For* $n = 1, 2, \dots, \infty$, *let* $\mathscr{G}_t^n = \sigma + (\tilde{D}^n(s), \tilde{X}_1(s), \dots, \tilde{X}_{\tilde{D}^n(s)}(s))\colon s \leq t\}$. *Then* $\pi_{\tilde{D}^n(t), n}$ *is the conditional distribution of* $(S_1^n(t), \dots, S_{\tilde{D}^n(t)}^n(t))$ *given* $\mathscr{G}_t^n$. *In particular, given* $\tilde{D}^n(t)$, $(\tilde{X}_1(t), \dots, \tilde{X}_{\tilde{D}^n(t)}(t))$ *is independent of* $(S_1^n(t), \dots, S_{\tilde{D}^n(t)}^n(t))$.

For $f \in \mathscr{D}(B)$ *with* $\|f\| < 1$, *define*

$$\begin{aligned} M_f(t) = {} & \prod_{i=1}^{\tilde{D}^n(t)} f(\tilde{X}_i(t)) \\ & - \int_0^{t \wedge \tau_1^n} \sum_{j=1}^{\tilde{D}^n(s)} \Biggl(\frac{Bf(\tilde{X}_j(s))}{f(\tilde{X}_j(s))} \\ & \qquad + I_{[2,n)}(\tilde{D}^n(s)) \frac{\tilde{D}^n(s) - 1}{2} \bigl(f(\tilde{X}_j(s)) - 1 \bigr) \Biggr) \\ & \qquad \times \prod_{i=1}^{\tilde{D}^n(s)} f(\tilde{X}_i(s))\, ds, \end{aligned} \tag{3.12}$$

where we assume (in the case $n = \infty$) that infinite products are zero. Then M_f is a $\mathscr{G}_t^n$ martingale.

PROOF. Consider $n < \infty$. For a bounded function $f(k, \underset{\sim}{s}, \underset{\sim}{x})$, define

$$Pf(k, \underset{\sim}{x}) = \sum_{s \in \Pi_{k,n}} f(k, \underset{\sim}{s}, \underset{\sim}{x}) \pi_{k,n}(s). \tag{3.13}$$

By the assumptions on B, the closure of

$$\tilde{G}^n f(k, \underset{\sim}{s}, \underset{\sim}{x}) = \sum_{i=1}^{k} B_i f(k, \underset{\sim}{s}, \underset{\sim}{x}) \tag{3.14}$$

generates a Markov semigroup on some separating subspace of $B(\mathscr{S}^n)$. For $n < \infty$, G^n is a bounded perturbation of $\tilde{G}^n$, so the closure of G^n will generate a Markov semigroup on the same subspace. By the same observation, $\hat{G}^n$ generates a Markov semigroup on some separating subspace of $B(\hat{\mathscr{S}}^n)$. For $n < \infty$, a straightforward calculation gives

$$PG^n f(k, \underset{\sim}{x}) = \hat{G}^n Pf(k, \underset{\sim}{x}) \tag{3.15}$$

and it follows from results of Kurtz and Ocone (1988) that $(\tilde{D}^n, \tilde{X}_1, \dots, \tilde{X}_{\tilde{R}^n})$ is a Markov process with generator $\hat{G}^n$ stopped at time τ_1^n and that $\pi_{\tilde{D}^n, n}$ is the conditional distribution of $(S_1^n, \dots, S_{\tilde{D}^n}^n)$ given $(\tilde{D}^n, \tilde{X}_1, \dots, \tilde{X}_{\tilde{D}^n})$. The final assertion of the theorem then follows from the fact that the integrand on the right of (3.12) is just $\hat{G}^n f(\tilde{D}^n, \tilde{X}_1, \dots, \tilde{X}_{\tilde{D}^n})$ for $f(k, \underset{\sim}{x}) = \Pi_{i=1}^k f(x_i)$.

For $n = \infty$, the conclusion of the theorem follows from the fact that

$$\lim_{n \to \infty} \left(\tilde{D}^n, \tilde{X}_1, \dots, \tilde{X}_{\tilde{D}^n} \right) = \left(\tilde{D}^\infty, \tilde{X}_1, \dots, \tilde{X}_{\tilde{D}^\infty} \right). \qquad \square \tag{3.16}$$

In the next theorem, we see that the martingale property of (3.12) uniquely characterizes the distribution of $(\tilde{D}^n(t), \tilde{X}_1(t), \dots, \tilde{X}_{\tilde{D}^n(t)}(t))$. In Theorem 3.9, we show that the Markov chain embedded in this process is just the urn model considered by Ethier and Griffiths (1987) and Ethier and Kurtz (1992) that generates a random sample from a population whose distribution of types is given by a stationary Fleming–Viot process.

For $x, y \in E^m$, define x to be equivalent to y, if x is a permutation of y. Let E^{*m} be the collection of equivalence classes and for $x \in E^{*m}$ define $|x| = m$. Define $S^* = \{\infty\} \cup \bigcup_{m=2}^{\infty} E^{*m}$, where ∞ is defined to be the limit of any sequence $\{x_n\}$ in $\bigcup_{m=2}^{\infty} E^{*m}$ in which $|x_n| \to \infty$.

THEOREM 3.6. *For $f \in \mathscr{D}(B)$ with $\|f\| < 1$, let $h_f(x) = \Pi_{k=1}^{|x|} f(x_k)$ if $|x| < \infty$ and $h_f(\infty) = 0$. Define*

$$Hh_f(x) = \sum_{j=1}^{|x|} \left(\frac{Bf(x_j)}{f(x_j)} + \frac{(|x| - 1)}{2} \big(f(x_j) - 1 \big) \right) \prod_{k=1}^{|x|} f(x_k) \tag{3.17}$$

if $|x| < \infty$ and $Hh_f(\infty) = 0$. Then the martingale problem for H is well posed.

Proof. Let

$$H_0 h_f(x) = \frac{2}{|x|-1} \sum_{j=1}^{|x|} \frac{Bf(x_j)}{f(x_j)} \prod_{k=1}^{|x|} f(x_k) \tag{3.18}$$

and

$$H_1 h_f(x) = \sum_{j=1}^{|x|} \big(f(x_j) - 1\big) \prod_{k=1}^{|x|} f(x_k). \tag{3.19}$$

Let $U_1, U_2, \ldots$ be independent copies of U as defined at the beginning of Section 2. For $x \in E^{*n}$, Y defined by

$$Y(t) = \big(U_1(x_1, 2t/(n-1)), \ldots, U_n(x_n, 2t/(n-1))\big)$$

is a solution of the martingale problem for H_0. Let L be the closure of the linear span D_s of $D = \{h_f : f \in \mathscr{D}(B), \|f\| < 1\}$. Let $\{T(t)\}$ be the semigroup generated by B and define.

$$u_f(t,x) = \prod_{k=1}^{|x|} T(2t/(|x|-1)) f(x_k)$$

Then

$$\frac{d}{dt} u_f(t,x) = H_0 u_f(t,x) \tag{3.20}$$

and it follows by Proposition 1.3.4 of Ethier and Kurtz (1986) that the closure of the linear extension H_0^e of H_0 generates a strongly continuous, contraction semigroup on L. Since the closure of the linear extension H_1^e of H_1 defines a bounded linear operator on L, it follows that the closure of $H_0^e + H_1^e$ generates a strongly continuous contraction semigroup on L. By Theorem 4.4.1 and Proposition 4.10.1 of Ethier and Kurtz (1986), the martingale problem for $H_0 + H_1$ is well posed. Suppose that V is a solution of the martingale problem for H and define $\gamma = \lim_{n\to\infty} \inf\{t : |V(t)| = n\}$. Let

$$\int_0^{\beta(t)} \frac{|V(s)| - 1}{2} ds = t. \tag{3.21}$$

Then $\hat{V}$ defined by $\hat{V}(t) = V(\beta(t))$ is a solution of the martingale problem for $H_0 + H_1$, and uniqueness for the martingale problem for $H_0 + H_1$ implies that the distribution of $V(\cdot \wedge \gamma)$ is uniquely determined by the distribution of $V(0)$. For $0 < \alpha < 1$, M_α defined by

$$M_\alpha(t) = \alpha^{|V(t)|} - \int_0^t \frac{(\alpha-1)(|V(s)|-1)}{2} \alpha^{|V(s)|} ds \tag{3.22}$$

is a martingale and since the integrand is nonnegative, L_α defined by $L_\alpha(t) = \alpha^{|V(t)|}$ is a nonnegative supermartingale. Since $\gamma = \inf\{t : L_\alpha(t-) = 0\}$ it follows that, with probability 1, for all $t > \gamma$, $L_\alpha(t) = 0$ and hence that $V(t) = \infty$. Consequently, the distribution of V is determined by the distribution of $V(0)$, giving the desired uniqueness. □

The next theorem exhibits the structure of stationary sampling distributions (equivalently moment measures) in the Fleming–Viot process. In particular, the theorem implies that one can obtain a realization of a sample of size n from the appropriate Fleming–Viot process at stationarity by the following algorithm:

1. Choose a realization of the genealogy of a sample of size n from the n-coalescent.
2. Independently, choose the type of the common ancestor of the sample from the stationary distribution of the mutation process.
3. Conditional on steps 1 and 2, follow the evolution of the types forward through the genealogy from the common ancestor. These types evolve according to the mutation process independently on distinct branches of the genealogical tree.

THEOREM 3.7. *Suppose the Fleming–Viot process is stationary.*

(a) $\tilde{D}^n(0) = 2$ *a.s. and the distribution of* $(\tilde{X}_1(0), \tilde{X}_2(0))$ *is that of* (X, X), *where the distribution of* $X \in E$ *is the stationary distribution of the mutation process. Further*, $(\tilde{X}_1(0), \tilde{X}_2(0))$ *is independent of* $\tilde{R}^n$ *on* $[0, \tau_1^n)$.

(b) *The distribution of* $(\tilde{X}_1(\tau_1^n -), \dots, \tilde{X}_n(\tau_1^n -))$ *is that of an independent sample of size n from the Fleming–Viot measure.*

PROOF. Note that $\tilde{D}^n(0) = |R^n(\tau_1^n -)| = 2$ a.s. Further, by right continuity of the mutation process, $\tilde{X}_1(0) = \tilde{X}_2(0) = X_1(-\tau_1^n)$, the type of particle 1 at the time $-\tau_1^n$ at which the first n particles at time zero first have a common ancestor. Now the distribution of $X_1(-\tau_1^n)$ depends on the mutation process for particle 1 on $(-\infty, -\tau_1^n]$. The value of τ_1^n, and indeed the behavior of $\tilde{R}^n$, depends only on the Poisson processes of look-downs from each of the first n levels, which is independent of the mutation process on the bottom level. Stationarity ensures that, for any t, the distribution of $X_1(t)$ is that of the stationary distribution of the mutation process and (a) follows.

For (b), observe that by construction $(\tilde{X}(\tau_1^n -), \tilde{X}_2(\tau_1^n -), \dots, \tilde{X}_n(\tau_1^n -))$ is a permutation of $(X_1(0), \dots, X_n(0))$ with the permutation independent of the values of $(X_1(0), \dots, X_n(0))$. □

The next theorem considers the *ancestral type measure*, that is, the frequency distribution of the types of the ancestors at a time $s < t$ of the individuals alive at time t.

THEOREM 3.8. *For* $s < t$, *let*

$$Z_t(s) = \frac{1}{|\Gamma(s, t)|} \sum_{k \in \Gamma(s, t)} \delta_{X_k(s)} \tag{3.23}$$

and let q be as in Theorem 2.4. *Then* $\lim_{s \to t} q(Z_t(s), Z(t)) = 0$ *a.s.*

PROOF. Since $\Gamma(s, t)$ depends only on $\{N_{ij}\}$ in the time interval $(s, t]$, $\Gamma(s, t]$ is independent of $X(s)$. Consequently, since the $\{X_k(s)\}$ are conditionally iid with distribution $Z(s)$, by Lemma A3.1, for any $f \in B(E)$ and $\varepsilon > 0$, there exist C and $\eta > 0$ depending only on $\|f\|$ and ε such that

$$P\left\{\left|\int f(x) Z_t(s, dx) - \int f(x) Z(s, dx)\right| > \varepsilon \middle| \Gamma(s,t)\right\} \le Ce^{-\eta|\Gamma(s,t)|}. \tag{3.24}$$

Furthermore, if we define $\tau_k = \sup\{s : |\Gamma(s,t)| = k\}$, $\Gamma(\tau_k -, t)$ is independent of $X(\tau_k -)$ and the sequence $(X_1(\tau_k -), X_2(\tau_k -), \dots)$ is exchangeable. (To check the exchangeability, approximate τ_k by $\tau_k^\delta = \max\{\delta m : m \in \mathbb{N}, |\Gamma(\delta m, t)| \le k\}$.) We then have

$$P\left\{\left|\int f(x) Z_t(\tau_k -, dx) - \int f(x) Z(\tau_k, dx)\right| > \varepsilon\right\} \le Ce^{-\eta k}. \tag{3.25}$$

For $f \in \mathscr{D}(B)$, the continuity of $\int f(x) Z(s, dx)$ and (3.25) imply that $\lim_{k\to\infty} \int f(x) Z_t(\tau_k -, dx) = \int f(x) Z(t, dx)$ a.s. Since $\Gamma(\tau_k -, t)$ and $\Gamma(\tau_k, t)$ differ by only one value,

$$\lim_{k\to\infty} \int f(x) Z_t(\tau_k, dx) = \int f(x) Z(t, dx) \quad \text{a.s.} \tag{3.26}$$

as well. As in (2.31),

$$\begin{aligned} P&\left\{\sup_{\tau_k \le s < \tau_{k+1}} \left|\int f(x) Z_t(s, dx) - \int f(x) Z_t(\tau_k, dx)\right| > \varepsilon\right\} \\ &\le P\left\{\sup_{\tau_k \le s < \tau_{k+1}} \left|\int f(x) Z_t(s, dx) - \int f(x) Z_t(\tau_k, dx) \right.\right. \\ &\qquad\qquad \left.\left. - \int_{\tau_k}^{s} \int Bf(x) Z_t(s, dx)\, ds\right| > \frac{\varepsilon}{2}\right\} \\ &\quad + P\left\{\int_{\tau_k}^{\tau_{k+1}} \int |Bf(x)| Z_t(s, dx)\, ds > \frac{\varepsilon}{2}\right\} \end{aligned} \tag{3.27}$$

and the left-hand side can be bounded by $Ce^{-\eta(k+1)}$, where C and η depend only on $\|f\|$, $\|Bf\|$ and ε. This estimate along with (3.26) confirms the desired convergence. □

Now suppose that the mutation operator has the form

$$Bf(x) = \frac{\theta}{2} \int (f(y) - f(x)) \eta(x, dy) \tag{3.28}$$

with stationary distribution π. Define $S = \bigcup_{n=2}^{\infty} E_n$ and, for clarity, denote an element of S by both its dimension and components (n, x). Consider the following discrete-time urn-type process, $\{U_k, k = 0, 1, 2, \dots\}$. Take $U_0 = (2, X_1, X_2)$, where $X_1 = X_2$ have distribution π. To obtain U_{k+1} from $U_k = (n, X_1, \dots, X_n)$:

1. For $i = 1, \dots, n$, with probability $(n-1)/(n(n+\theta-1))$ duplicate X_i to obtain $U_{k+1} = (n+1, X_1, \dots, X_n, X_i)$.

2. For $i = 1, \dots, n$, with probability $\theta/(n(n+\theta-1))$ mutate X_i to obtain $U_{k+1} = (n, X_1, \dots, X_{i-1}, Y, X_{i+1}, \dots, X_n)$, where the distribution of Y is given by $\eta(X_i, \cdot)$.

Write $\sigma_n = \min\{k: U_k \in n \times E^n\}$ for th epoch at which U first has n components. The next result was first obtained by Ethier and Griffiths (1987).

THEOREM 3.9. *For the urn model U just described, the joint distribution of the n types in the urn immediately before the time σ_{n+1} is that of a sample of size n from the Fleming–Viot process at stationarity.*

PROOF. Observe that the urn model U is simply the embedded Markov chain of the process $(\tilde{D}^\infty, \tilde{X}_1, \dots, \tilde{X}_{\tilde{D}^\infty})$. Further, $\tilde{D}^\infty(0) = 2$ a.s. and $\tilde{X}_1(0) = \tilde{X}_2(0)$ have distribution π exactly as in the proof of Theorem 3.7. If we write $\tilde{\tau}_n = \inf\{t: \tilde{D}^\infty(t) = n\}$, then the dynamics of $(\tilde{D}^\infty, \tilde{X}_1, \dots, \tilde{X}_{\tilde{D}^\infty})$ on $(0, \tilde{\tau}_{n+1})$ are the same as those of $(\tilde{D}^n, \tilde{X}_1, \dots, \tilde{X}_{\tilde{D}^n})$ on $(0, \tau_1^n)$, so the result follows from Theorem 3.7. □

Finally, we remark that Theorem 3.8 ensures that the empirical measure associated with the collection of types in the urn process converges as $k \to \infty$ and that the distribution of the limiting empirical measure is the stationary distribution for the Fleming–Viot process. We study more detailed properties of these urn processes in Donnelly and Kurtz (1996).

4. Ergodicity and convergence to stationarity.

The particle construction provides a useful device for coupling Fleming–Viot processes. We illustrate this device with results on rate of convergence to stationarity. Recall that the total variation distance between two probability measures ν_1 and ν_2 on a space S is given by

$$\|\nu_1 - \nu_2\| = \sup_{\Gamma \in \mathscr{B}(S)} |\nu_1(\Gamma) - \nu_2(\Gamma)|. \tag{4.1}$$

THEOREM 4.1. *Let $\{T^*(t)\}$ denote the adjoint mutation semigroup on $\mathscr{P}(E)$ and let $\{\mathbb{S}^*(t)\}$ denote the adjoint Fleming–Viot semigroup on $\mathscr{P}(\mathscr{P}(E))$. Suppose that the mutation process is strongly ergodic with stationary distribution π in the sense that*

$$\lim_{t \to \infty} \|T^*(t)\nu - \pi\| = 0 \tag{4.2}$$

for all $\nu \in \mathscr{P}(E)$. Then the Fleming–Viot process is strongly ergodic with a stationary distribution Π, and for each $\Psi \in \mathscr{P}(\mathscr{P}(E))$ there exist independent random variables T_m and T_d such that

$$\|\mathbb{S}^*(t)\Psi - \Pi\| \le P\{T_m + T_d > t\}, \tag{4.3}$$

$$P\{T_m > t\} = \|T^*(t)\nu - \pi\| \tag{4.4}$$

for $\nu = \int \mu \Psi(d\mu)$ *and*

$$T_d =_{\mathscr{D}} \sum_{i=2}^{\infty} E_i, \tag{4.5}$$

where the E_i *are independent exponential random variables with* $E[E_i] = \binom{i}{2}^{-1}$, $i = 2, 3, \dots$. *In particular, strong ergodicity of the mutation process guarantees strong ergodicity of the Fleming–Viot process.*

REMARK 4.2. A different coupling of Fleming–Viot processes and proof that strong ergodicity of the mutation process implies strong ergodicity of the associated Fleming–Viot process is given in Ethier and Kurtz (1995).

PROOF OF THEOREM 4.1. We will couple two versions of the particle process, X and $\tilde{X}$, with $\tilde{X}$ stationary. Then

$$\|\mathbb{S}^*(t)\Psi - \Pi\| \le P\{X \ne \tilde{X}\}. \tag{4.6}$$

Make the demography identical in each process. That is, for each $i < j$, use the same Poisson process of look-downs between levels j and i in the two processes. Further, recalling the construction defined at the beginning of Section 2, use the same mutation processes $\{U_{ijk},\ 1 \le i < j,\ 1 \le k < \infty\}$ and $\{U_{i0},\ i \ge 2\}$ in both X and $\tilde{X}$. Let $X(0) = (X_1(0), X_2(0), \dots)$ be exchangeable with de Finetti measure determined by Ψ and $\tilde{X}(0) = (\tilde{X}_1(0), \tilde{X}_2(0), \dots)$ exchangeable with de Finetti measure determined by Π. There exists a maximal coupling between the mutation process with initial distribution ν and the mutation process with initial distribution π [see Sverchkov and Smirnov (1990)], and by an appropriate pasting argument [see Ethier and Kurtz (1986), Lemma 4.5.15] we can construct X_1 and $\tilde{X}_1$ with coupling time T_m so that

$$P\{X_1(t) \ne \tilde{X}_1(t)\} = P\{T_m > t\} = \|T^*(t)\nu - \pi\|. \tag{4.7}$$

Let $\Gamma(s,t)$ be as in Section 3 and, for each t, let $\{T_t = \inf\{s: |\Gamma(t-s,t)| = 1\}$. Since the demography is independent of X_1 and $\tilde{X}_1$, T_t is independent of T_m. Since $X_1(t) = \tilde{X}_1(t)$, after time T_m, it follows from the construction that

$$P\{X(t) \ne \tilde{X}(t)\} \le P\{t - T_t > T_m\} \tag{4.8}$$

and since T_t is distributed as the appropriate sum of exponentials, the result follows. □

We note that similar arguments may be used to study rates of convergence for various functionals of the Fleming–Viot process. In particular, convergence of sampling distributions is related, through $X_1(t), \dots, X_n(t)$, to the behavior of the n-coalescent.

For certain kinds of mutation processes, a slightly different construction is also informative.

THEOREM 4.3. *Suppose the mutation operator has the form*

$$Bf(x) = \frac{\theta}{2}\int (f(y) - f(x))\,\nu_0(dy) \tag{4.9}$$

for $\theta > 0$. *Then*

$$\|\mathbb{S}^*(t)\Psi - \Pi\| \geq P\{\tilde{T} > t\}, \tag{4.10}$$

where $\tilde{T} = \mathscr{D}\sum_{i=1}^{\infty}\tilde{E}_i$ *in which the* $\tilde{E}_i$ *are independent exponential random variables with* $E[\tilde{E}_i] = 2/(i(i+\theta-1))$, $i = 1, 2, \ldots$. *Further, equality holds in* (4.10) *if* $\nu = \int \mu\Psi(d\mu)$ *and* ν_0 *are mutually singular.*

REMARK 4.4. This theorem is originally due to Ethier and Griffiths (1993), who derived it directly from the transition density of the Fleming–Viot process. The construction given here and properties of the lines of descent and associated genealogical processes [Donnelly and Tavaré (1987a)] can be used to give an alternative derivation of the transition density obtained originally by Ethier and Griffiths (1993).

PROOF OF THEOREM 4.3. Again, we construct two versions of the particle process, X and $\tilde{X}$. Use the same demography in each as before. In addition, construct the mutation processes in each from a sequence K_i^{θ}, $i = 1, 2, \ldots,$ of independent Poisson processes on $[0, \infty) \times E$, each with mean measure density $(\theta/2)\,dt \times \nu_0(dx)$. Think of the first coordinate of K_i^{θ} as giving the times at which mutations occur at level i (in each process) and the second coordinate as specifying the type at level i immediately after the associated mutation. Then $K_i^{\theta}((a, b] \times E)$ counts the number of mutations on level i in the time interval $(a, b]$.

Now consider a slightly different genealogical process which is known [Griffiths (1980)] as the lines of descent process. Fix $t > 0$ and, for each particle at time t, trace the ancestry backward in time. In addition to having the level of an ancestor jump each time it encounters a look-down, "kill" the ancestor each time it encounters a mutation. Formally, recall that for $s \leq t$, $a_j(s, t)$ is the level at time $t - s$ of the ancestor of particle j at time t. Write $\sigma_{j1} < \sigma_{j2} < \cdots < \sigma_{jk_j}$ for the jump times of $a_j(\cdot, t)$ over $[0, s]$, and write a_{jl} for the value of $a_j(\cdot, t)$ on $(\sigma_{jl-1}, \sigma_{jl}]$, $l = 1, 2, \ldots, k_j + 1$, with $\sigma_{j0} \equiv 0$ and $\sigma_{jk_j+1} \equiv s$. Define

$$\tilde{a}_j(s,t) = \begin{cases} a_j(s,t), & \text{if } \sum_{l=1}^{k_j+1} K_{a_{jl}}^{\theta}\big([\sigma_{jl-1}, \sigma_{jl}] \times E\big) = 0, \\ \delta, & \text{otherwise.} \end{cases} \tag{4.11}$$

Note that $\tilde{a}_j(s, t] = \delta$ if there has been a mutation since time s in the ancestral line of the jth particle at time t.

For our purposes, it is enough to study $\tilde{\Gamma}(s, t) = \{\tilde{a}_j(s, t), j = 1, 2, \ldots, \tilde{a}_j(s, t) \neq \delta\}$, the levels of "surviving" ancestors at time s of particles at time t. For a detailed description of an associated equivalence relation, or frequency-valued process, see Donnelly and Tavaré (1987a, b). Note that on $[0, t]$, $L(s) = |\tilde{\Gamma}((t-s)-, t)|$ is a Markov death process on $\{0, 1, \dots, \infty\}$ with death rates $\frac{1}{2}k(k+\theta-1)$ from state k, $k = 1, 2, \dots$, and $L(0) = \infty$. That is, there is a "death" if there is either a look-down between two of the ancestors $\left[\text{rate } \binom{k}{2}\right]$ or a mutation in one of the ancestors (rate $k\theta/2$).

Since $X_i(t)$ is the type of the most recent mutation in the ancestral line of the ith level particle at time t and since the coupled processes X and $\tilde{X}$ are constructed using the same demography and mutation process, $\tilde{a}_i(0, t) = \delta$ implies $X_i(t) = \tilde{X}_i(t)$. Thus $X = \tilde{X}$ on $\{L(t) = 0\}$ and we have

$$P\{Z(t) = \tilde{Z}(t)\} \geq P\{X(t) = \tilde{X}(t)\} \geq P\{L(t) = 0\} \tag{4.12}$$

giving the desired inequality. If ν and ν_0 are mutually singular, then $L(t) > 0$ implies some level of X has its original type which, by Lemma 5.1 below, must correspond to an atom of $Z(t)$. Consequently, $P\{Z(t) \neq \tilde{Z}(t)\} \geq P\{L(t) > 0\}$ implying equality in (4.12) and (4.10). □

5. Sample path properties. In this section we describe a variety of sample-path properties for the Fleming–Viot process and their relationship to the particle model. If the mutation operator for the process is of the form

$$Bf(x) = \frac{\theta}{2} \int_E (f(y) - f(x)) \eta(x, dy) \tag{5.1}$$

we will say that the process has mutation rate θ. If η is nonatomic for all x (i.e., all mutants are new), then we will say that the process is an infinite-alleles model with mutation rate θ. Note that in this case the values of the Fleming–Viot process are purely atomic and the descending order statistics of the sizes of the atoms form an infinite-alleles diffusion model as in Ethier and Kurtz (1981). [See Ethier and Kurtz (1987).] The particle model with mutation rate θ can be constructed by associating with each level i a Poisson process K_i^θ with intensity $\theta/2$, where the K_i^θ are independent and independent of the N_{ij} determining the look-downs. The K_i^θ then give the times at which mutations occur. In particular, if τ is a jump time for K_i^θ, then $P\{X_i(\tau) \in G | \mathscr{F}_{\tau-}\} = \eta(X_i(\tau-), G)$.

In the next theorem we extend the assertion of purely atomic values to any Fleming–Viot process with a pure jump mutation process. We will need the following lemma, which describes the behavior of the atom in the Fleming–Viot process that corresponds to a mutant x that is "stable" in the sense that the mutation process has a nonzero exponentially distributed holding time in x. If the mutation process is pure jump, then (by definition) all states are stable; however, mutation processes that are not pure jump may have some stable states.

For $s < t$, let $a_j(s, t)$ be as in Section 3 and define $X_j(s, t) = X_{a_j(s,t)}(s)$. That is, $X_j(s, t)$ gives the type of the ancestor at time s of the jth level particle at time t.

LEMMA 5.1. *Suppose that a mutation occurs at time* τ_1 *in level* k *of the particle model and the mutant type* x_0 *is stable with holding time parameter* $\lambda \equiv \lambda(x_0)$. *Let* τ_2 *be the time of the next mutation or look-down on level* k *and set* $\Delta = \tau_2 - \tau_1$. (Δ *will be exponentially distributed with parameter* $\lambda + k - 1$.) *Let* $\xi_j(t) = 1$ *if there exists* $s \geq 0$ *such that* $a_j(\tau_1 + s, \tau_1 + t) = k$ *and* $X_j(\tau_1 + u, \tau_1 + t) = X_j(\tau_1 + s, \tau_1 + t)$ *for* $s \leq u \leq t$ *and* $\xi_j(t) = 0$ *otherwise. Then the limit*

$$S(t) = \lim_{m \to \infty} \frac{1}{m} \sum_{j=1}^{m} \xi_j(t) \tag{5.2}$$

exists a.s. Note that $S(t)$ *is the fraction of the levels at time* $\tau_1 + t$ *that are of type* x_0 *by direct descent from the mutant on level* k. *Further, up to time* Δ, S *is a Wright–Fisher diffusion with generator*

$$Gf(s) = \tfrac{1}{2}s(1-s)f''(s) + (1 - (\lambda + k)s)f'(s) \tag{5.3}$$

and $S(0) = 0$. *In particular,* $S(t) > 0$ *for* $0 < t \leq \Delta$.

PROOF. Clearly the distribution of S will depend only on λ and not on other properties of the mutation process, so without loss of generality, we can assume that the process is an infinite-alleles model with type space $E = [0, 1)$ and mutation operator

$$Bf(x) = \lambda \int_0^1 (f(y) - f(x))\, dy. \tag{5.4}$$

To simplify notation, assume $\tau_1 = 0$. (Otherwise translate the Poisson processes $\{N_{ij}\}$ and $K_i^{2\lambda}\}$ in the construction of the particle model by the stopping time τ_1.) Define $\tilde{N}_{ij} = N_{i+k, j+k}$ for $1 \leq i < j$, and define $\tilde{K}_i = K_{i+k}^{2\lambda} + \sum_{j=1}^{k} N_{j(i+k)}$ for $i \geq 1$. Note that $\tilde{K}_i$ is Poisson with parameter $\lambda + k$. We define a new particle model whose demography is determined by $\{\tilde{N}_{ij}\}$ and whose mutation times are determined by $\{\tilde{K}_i\}$. Level i in the new model is level $k + i$ in the old. The type at new level i mutates at the jump times of $K_{i+k}^{2\lambda}$ as before and, in addition, for $1 \leq j \leq k$, at the jump times of N_{jk+i}, the type at new level i is set to a new type j. The new model then is a particle model as constructed in Section 2 with type space $\tilde{E} = [0, 1) \cup \{1, \dots, k\}$ and mutation operator

$$\tilde{B}f(x) = \sum_{j=1}^{k} (f(j) - f(x)) + \lambda \int_0^1 (f(y) - f(x))\, dy. \tag{5.5}$$

Note that the type at level i at time t in the new model is type k if and only if in the original model there exists $s > 0$ such that $a_{k+i}(s, t) = k$ and $X_{k+i}(u, t) = X_{k+i}(t)$ for $s \leq u \leq t$. That is, the type of the particle at level i at time t in the new model is k if and only if $\xi_i(t) = 1$. [Since in the new model, mutation to type $j \leq k$ corresponds to a look-down to level j in the original model, the assumption that $a_{k+i}(s, t) = k$ rules out the possibility that $\tilde{X}_i(u, t) = j$ for any $s \leq u \leq t$ in the new model.] Consequently, the limit in

(5.2) exists by exchangeability in the new model, and up to time Δ, $S(t)$ is just $A_k(t)$, the size of the atom at $k \in \tilde{E}$ in the new model. Taking $g = I_{\{k\}}$ in Lemma 2.3, we see that A_k is a solution of the martingale problem for G, and the first part of the lemma follows. To see that $A_k(t) > 0$ for all $t > 0$, check that the solution of $Gf_0 = 0$ is unbounded at 0 and that the solution of $Gf_1 = 1$ is continuous on $[0, 1]$. These calculations imply that the solution of the martingale problem with initial value 0 leaves 0 instantaneously and that no solution hits 0 at a positive time. □

THEOREM 5.2. *Let the mutation operator be of the form*

$$Bf(x) = \lambda(x) \int_E (f(y) - f(x)) \eta(x, dy) \tag{5.6}$$

and suppose that the mutation process has the property that

$$\sup_{0 \le s \le t} \lambda(X_1(s)) < \infty \quad a.s. \text{ for each } t > 0.$$

Then $P\{Z(t) \in \mathscr{P}_a(E),\ t > 0\} = 1$. [$\mathscr{P}_a(E)$ *denotes the collection of purely atomic probability measures on* E.]

PROOF. The condition on the mutation process ensures that X_k is a pure jump process for each k, that is, there is no accumulation point of discontinuities on any level. Furthermore, the independence assumptions on mutations between look-downs ensures that no two levels have simultaneous discontinuities. Fix a level k and let $[a, b)$ be a time interval on which X_k is constant and let $\lambda_0 = \lambda(X_k(a))$. By Lemma 5.1, there is a strictly positive atom in $Z(t)$ at $X_k(a)$ for all $a < t \le b$. On each level k, there are countably many time intervals $[\alpha_{ki}, \alpha_{k(i+1)})$ on which X_k is constant with $\alpha_{k0} = 0 < \alpha_{k1} < \alpha_{k2} < \cdots$. Consequently, with probability 1 for all i, k, $Z(t, \{X_k(\alpha_{ki})\}) > 0$ for all $t \in (\alpha_{ki}, \alpha_{k(i+1)})$. Since with probability 1, $\alpha_{ki} \ne \alpha_{mj}$ for $(m, j) \ne (k, i)$, for all $t > 0$, at most one level fails to correspond to an atom of $Z(t)$, and the theorem follows. □

THEOREM 5.3. *Assume that B is of the form* (5.1) *and define μ by*

$$\mu(x, F) = \sum_{k=1}^{\infty} \frac{1}{2^k} \eta^k(x, F), \tag{5.7}$$

where $\eta^1(x, F) = \eta(x, F)$ and $\eta^{k+1}(x, F) = \int \eta^k(y, F)\eta(x, dy)$. Let S_x be the support of the measure $\mu(x, \cdot)$. That is, $S_x = \{y\colon \mu(x, B_\varepsilon(y)) > 0$ for all $\varepsilon > 0\}$ where $B_\varepsilon(y)$ is the ball of radius ε centered at y. Let C_t denote the closure in E of the set $\{X_i(t), i \ge 1\}$. Fix $t > 0$. Suppose there exists k and $h > 0$ such that $X_k(s) = x_0$ for $t - h \le s < t$. Then with probability 1, $C_t \supset S_{x_0}$. *In particular, if the mutation process has generator of the form* (5.1) *and is recurrent in the sense that for any choice of $X_1(0)$, X_1 visits every open ball infinitely often* (*hence $S_x = E$ for every x*), *then with probability* 1, $C_t = E$.

REMARK 5.4. Note that this result is a probability 1 statement for each fixed t. The next results will show that even in the recurrent case, in general there exist t such that $C_t \neq E$.

PROOF OF THEOREM 5.3. Proposition 2.2 of Donnelly and Kurtz (1996) ensures that all but finitely many of the types that appear in the mutation process on the genealogical tree for the population at time t will be represented in the population at time t. Lemma 5.1 ensures that the population at time t will include infinitely many replicates of x_0. The results on the urn model in Donnelly and Kurtz (1995) then ensure that the population at time t will include infinitely many kth order mutants from x_0. In particular, it will include an infinite independent sample from the distribution $\eta^k(x_0, \cdot)$, and hence, for any $y \in E$ and any $\varepsilon > 0$ such that $\mu(B_\varepsilon(y)) > 0$, $B_\varepsilon(y)$ will contain infinitely many of the $X_i(t)$. □

THEOREM 5.5. *For an infinite-alleles model, with probability* 1, *there will exist times t at which the Fleming–Viot measure consists of a single atom if and only if* $\theta < 1$. *If* $\theta \geq 1$, *then there will always be an infinite number of atoms.*

REMARK 5.6. Schmuland (1991) proves this result using Dirichlet forms.

PROOF OF THEOREM 5.5. Let $S_1(t)$ denote the size of the largest atom in $Z(t)$ and define $\tau_1 = \inf\{t: S_1(t) = 1\}$. Define recursively $\alpha_1 = \inf\{t: S_1(t) \geq \frac{3}{4}\}$, $\beta_k = \inf\{t > \alpha_k: S_1(t) \leq \frac{1}{2}\}$ and $\alpha_{k+1} = \inf\{t > \beta_k: S_1(t) \geq \frac{3}{4}\}$. Fix a time interval $[\alpha_k, \beta_k)$ and define $\tilde{S}(t) = S_1(\alpha_k + t)$. By the strong Markov property, we can let g in Lemma 2.3 be the indicator of the location of the largest atom at time α_k. Noting that this location does not change during the time interval, Lemma 2.3 implies that $\tilde{S}$ satisfies

$$\tilde{S}(t) = S_1(\alpha_k) + \int_0^t \sqrt{\tilde{S}(s) - \tilde{S}^2(s)}\, dW_k(s) - \int_0^t \frac{\theta}{2}\tilde{S}(s)\, ds, \qquad t < \beta_k - \alpha_k,$$

for some standard Brownian motion W_k. Consequently, on any such time interval S_1 behaves like the Wright–Fisher diffusion with mutation away from the allele under consideration at rate θ, that is, with generator

$$G_\theta f(s) = \frac{1}{2}s(1-s)f''(s) - \frac{\theta}{2}sf'(s). \tag{5.8}$$

For this diffusion, the boundary at 1 is accessible; that is, S_1 can reach 1 if and only if $\theta < 1$. If $\theta < 1$, a renewal argument shows that $\tau_1 < \infty$. If $\theta \geq 1$, $\tau_1 > \beta_k$ for every k. Since $\{\beta_k - \alpha_k\}$ is iid by the strong Markov property for S_1, we must have $\lim_{k\to\infty} \beta_k = \infty$ and hence that $\tau_1 = \infty$.

Now assume that $\theta \geq 1$. Let $S_2(t)$ denote the sum of the sizes of the two largest atoms at time t and define $\tau_2 = \inf\{t: S_2(t) = 1\}$. Let $\sigma_n = \inf\{t: S_1(t) \geq 1 - 1/4^n\}$, and note that $\lim_{n\to\infty} \sigma_n = \infty$, since S_1 never equals 1. Fix n and define $\alpha_1 = \inf\{t: S_2(t) \geq 1 - 1/4^{n+1}\}$, $\beta_k = \inf\{t > \alpha_k: S_2(t) \leq$

$1 - 1/(2 \cdot 4^n)\}$ and $\alpha_{k+1} = \inf\{t > \beta_k : S_2(t) \geq 1 - 1/4^{n+1}\}$. If $\beta_k < \sigma_n$, then, as above, on the interval $[\alpha_k, \beta_k)$, the locations of the two largest atoms do not change and S_2 behaves like the Wright–Fisher diffusion with generator G_θ. Consequently, it follows that $\tau_2 \geq \beta_k \wedge \sigma_n$, but by a renewal argument, $\beta_k \geq \sigma_n$ for some k. Since $\sigma_n \to \infty$, it follows that $\tau_2 = \infty$ and hence that there will always be at least three atoms. An induction argument now shows that there will always be at least k atoms, and the final assertion of the theorem follows. □

COROLLARY 5.7. *In the infinite-alleles particle model with mutation rate* θ, *with probability* 1, *there will exist times at which all particles are of the same type if and only if* $\theta < 1$.

PROOF. Let τ be the first time that the Fleming–Viot measure consists of a single atom. Then by Corollary 2.6 the collection of particle types $\{X_k(\tau)\}$ is exchangeable with de Finetti distribution $Z(\tau)$, and hence with probability 1, all $X_k(\tau)$ are the same. Since the empirical measure

$$Z_m(t) = \frac{1}{m} \sum_{k=1}^{m} \delta_{X_k(t)} \tag{5.9}$$

converges to Z uniformly on bounded time intervals, all $X_k(t)$ will be the same only if $Z(t)$ consists of a single atom. However, by Theorem 5.5, there exists a t at which $Z(t)$ consists of a single atom if and only if $\theta < 1$. □

THEOREM 5.8. *Let* $Bf(x) = \lambda(x)\int_E (f(y) - f(x))\eta(x, dy)$ *with* $\bar{\lambda} = \sup_x \lambda(x) < \infty$.

(a) *If* $\lambda(x) < \frac{1}{2}$ *and for every* n *there exists* $t > n$ *such that* $X_1(t) = x$ (*in particular*, *if* X_1 *is stationary and ergodic and the stationary distribution has an atom at* x), *then*, *with probability* 1, *there exist times* t *such that* $Z(t) = \delta_x$. *If* $\lambda(x) \geq \frac{1}{2}$, *there exist no such times*.

(b) *For* $A \subset E$, *let*

$$\bar{\lambda}_A \equiv \sup_{x \in A} \lambda(x)\eta(x, A^c) \quad \text{and} \quad \underline{\lambda}_A \equiv \inf_{x \in A} \lambda(x)\eta(x, A^c).$$

If $\bar{\lambda}_A < \frac{1}{2}$ *and for every* n *there exists* $t > n$ *such that* $X_1(t) \in A$ [*in particular*, *if* X_1 *is stationary and ergodic and the stationary distribution* π *for the mutation process satisfies* $\pi(A) > 0$], *then*, *with probability* 1, *there exist times* t *such that* $Z(t, A) = 1$. *If* $\underline{\lambda}_A \geq \frac{1}{2}$, *there exist no such times*.

(c) *Let* $A_1, A_2, \ldots \subset E$ *with* $\lambda \equiv \sup_k \sup_{x \in A_k} \lambda(x)\eta(x, A_k^c) < \frac{1}{2}$. *If for every* n *there exists* $t > n$ *such that* $X_1(t) \in \bigcup A_k$, *then*, *with probability* 1, *there exist times* t *such that for some* k, $Z(t, A_k) = 1$. [*For example*, *if* E *is countable and* B *corresponds to a transient Markov chain with* $\sup_x \lambda(x) < \frac{1}{2}$, *then there will exist times* t *such that* $Z(t)$ *has a single atom*.]

REMARK 5.9. (a) Similar results have been derived independently in Overbeck, Röckner and Schmuland (1995) using Dirichlet form techniques.

(b) The existence of times at which $Z(t, A) = 1$ in (b) or $Z(t, A_k) = 1$ in (c) does not depend on the boundedness or the form of the generator off the set A (resp. $\bigcup_k A_k$).

PROOF OF THEOREM 5.8. Part (a) is a special case of part (b). Assume $\bar{\lambda}_A < \frac{1}{2}$. Let $[\tau_1, \tau_2)$ be an interval on which X_1 is in A. (We can assume that the length of the interval is bounded below by an exponential random variable Δ with parameter $\bar{\lambda}_A$.) For any bounded measurable g, by Corollary 2.8, we can write $\langle g, Z(t)\rangle$ as

$$\begin{aligned}\langle g, Z(t)\rangle = \langle g, Z(0)\rangle &+ \int_0^t \sqrt{\langle g^2, Z(s)\rangle - \langle g, Z(s)\rangle^2}\, dW(s)\\ &+ \int_0^t (\langle Bg, Z(s)\rangle + g(X_1(s)) - \langle g, Z(s)\rangle)\, ds.\end{aligned} \tag{5.10}$$

Assume for simplicity that $\tau_1 = 0$. Then taking $g = I_A$, for $t \le \tau_2$,

$$\begin{aligned}Z(t, A) &= Z(0, A) + \int_0^t \sqrt{Z(s, A) - Z(s, A)^2}\, dW(s)\\ &\qquad + \int_0^t (\langle BI_A, Z(s)\rangle + 1 - Z(s, A))\, ds\\ &\ge Z(0, A) + \int_0^t \sqrt{Z(s, A) - Z(s, A)^2}\, dW(s)\\ &\qquad + \int_0^t (-\langle \lambda(\cdot)\eta(\cdot, A^c) I_A, Z(s)\rangle + 1 - Z(s, A))\, ds\\ &\ge Z(0, A) + \int_0^t \sqrt{Z(s, A) - Z(s, A)^2}\, dW(s)\\ &\qquad + \int_0^t \left(-\bar{\lambda}_A Z(s, A) + 1 - Z(s, A)\right) ds.\end{aligned} \tag{5.11}$$

Note that $Z(t, A)$ cannot be identically zero on any interval $[0, \varepsilon]$. [Replace $Z(s, A)$ by zero on the right side of (5.11) and observe the contradiction.] If $Z(\varepsilon, A) > 0$ and $\varepsilon < \Delta$, then let V satisfy

$$\begin{aligned}V(t) = &\int_\varepsilon^t \sqrt{V(s) - V(s)^2}\, dW(s)\\ &+ \int_\varepsilon^t \left(-\bar{\lambda}_A V(s) + 1 - V(s)\right) ds.\end{aligned} \tag{5.12}$$

Then by Lemma A4.1, $Z(t, A) > V(t)$ at least until the minimum of $\beta = \inf\{t\colon V(t) = 0 \text{ or } 1\}$ and Δ. It is easy to check that $V(t) > 0$ for all $t > \varepsilon$ and that $\beta \equiv \inf\{t\colon V(t) = 1\} < \infty$. Since β is independent of Δ, $\rho = P\{\beta < \Delta\} > 0$, and since there are infinitely many disjoint intervals satisfying the conditions on $[\tau_1, \tau_2]$, it follows that $Z(t, A) = 1$ for a sequence of arbitrarily large values of t. Part (c) follows by the same argument.

The converse result in part (b) follows similarly by first observing that

$$
\begin{aligned}
Z(t,A) &= Z(0,A) \\
&\qquad + \int_0^t \sqrt{Z(s,A)-Z(s,A)^2}\, dW(s) \\
&\qquad + \int_0^t \bigl(\langle BI_A, Z(s)\rangle + 1 - Z(s,A)\bigr)\, ds \\
&= Z(0,A) + \int_0^t \sqrt{Z(s,A)-Z(s,A)^2}\, dW(s) \\
&\qquad + \int_0^t \bigl(\langle \lambda(\cdot)\eta(\cdot,A) I_{A^c}, Z(s)\rangle - \langle \lambda(\cdot)\eta(\cdot,A^c) I_A, Z(s)\rangle \\
&\qquad\qquad + 1 - Z(s,A)\bigr)\, ds \\
&\le Z(0,A) + \int_0^t \sqrt{Z(s,A)-Z(s,A)^2}\, dW(s) \\
&\qquad + \int_0^t \bigl(-\underline{\lambda}_A Z(s,A) + (1+\bar{\lambda})(1-Z(s,A))\bigr)\, ds. \qquad \square
\end{aligned}
\tag{5.11'}
$$

THEOREM 5.10. *Let* $\{N_{ij}\}$ *and* $\{K_i^\theta\}$ *be the Poisson processes determining the demography and times of mutation and, for each* t, *let* T_t *be the genealogical tree of the population at time* t *determined by the* $\{N_{ij}\}$.

(a) *With probability* 1, *at no time* t *will the genealogical tree* T_t *have finite total length.*

(b) *If* $\theta < 1$, *then, with probability* 1, *there exist times* t *such that* T_t *contains no jump times of the* $\{K_i^\theta\}$.

(c) *If* $\theta \ge 1$, *then, with probability* 1, *for every* t, T_t *will contain infinitely many jump times of the* $\{K_i^\theta\}$.

PROOF. For $\theta \ge 1$, the fact that in an infinite-alleles model $Z(t)$ must have infinitely many atoms for all t implies T_t must contain infinitely many mutation times, that is, jump times of the K_i^θ, proving (c). For $\theta < 1$, the fact that there exist t at which $Z(t)$ consists of a single atom implies there exist t for which T_t contains no jump times of the K_i^θ. If there existed a time t such that T_t had finite length, then by the independence of T_t and $\{K_i^\theta\}$, there would be positive probability of $Z(t)$ having only finitely many atoms. However, this is false for $\theta \ge 1$, so T_t must have infinite length for all t. $\square$

There is an alternative proof of Theorem 5.8 (and Theorem 5.11 below) employing coupling arguments and the properties of the genealogy given by Theorem 5.10. For example, the jump times of the mutation process out of the set A in Theorem 5.8(b) can be obtained by thinning a Poisson process with rate $\bar{\lambda}_A$. Theorem 5.10(b) guarantees that there will be an infinite number of times t for which the genealogical tree of the population at time t contains no jumps of this Poisson process, so that if the common ancestor has type in A, $Z(t, A) = 1$.

THEOREM 5.11 [Ohta–Kimura model (1973, 1974)]. *Let* $E = \mathbb{Z}$ *and*

$$Bf(k) = \lambda\left(\frac{f(k+1) + f(k-1)}{2} - f(k)\right). \tag{5.13}$$

(a) *If* $\lambda < \frac{1}{2}$, *then, with probability* 1, *there exist times* t *at which* $Z(t, \{0\}) = 1$.

(b) *If* $\frac{1}{2} \leq \lambda < 1$, *then, with probability* 1, *there exist times* t *at which* $Z(t, \{0, 1\}) = 1$, *but no time* t *at which* $Z(t, \{0\}) = 1$. [*In fact, with probability* 1, *there is no time* t *at which* $Z(t, \{k\colon k \text{ even}\}) = 1$.]

(c) *If* $\lambda \geq 1$, *then, with probability* 1, *there exist times* t *at which* $Z(t, \{-1, 0, 1\}) = 1$, *but no time* t *at which* $Z(t, \{0, 1\}) = 1$. [*In fact, there is no time* t *at which* $Z(t, \{k\colon k = 0 \text{ or } 1 \bmod 3\}) = 1$.]

PROOF. Part (a) is an immediate consequence of Theorem 5.8(a) and the recurrence of symmetric random walk. In part (b), the fact that $Z(t, \{0, 1\}) = 1$ for some t is an immediate consequence of Theorem 5.8(b) and the nonexistence of a time at which $Z(t, \{0\}) = 1$ follows from the fact that $Z(t, \{k\colon k \text{ even}\})$ is a Wright–Fisher diffusion with generator

$$G_b f(s) = \tfrac{1}{2}s(1-s)f''(s) + (\lambda(1-s) - \lambda s)f'(s). \tag{5.14}$$

Let $A = \{k\colon k = 0 \text{ or } 1 \bmod 3\})$. In part (c), the fact that there are no times at which $Z(t, A) = 1$ follows from the fact that $Z(t, A)$ is a Wright–Fisher diffusion with generator

$$G_c f(s) = \frac{1}{2}s(1-s)f''(s) + \left(\frac{\lambda}{2}(1-s) - \lambda s\right)f'(s). \tag{5.15}$$

To see that $Z(t, \{+1, 0, 1\}) = 1$ for some time t, note that the recurrence of X_1 and Lemma 1 ensure that for any $m > 2$, $\alpha_1 = \inf\{t\colon Z(t, \{0\}) \geq 1 - 1/m\lambda\}$, $\beta_k = \inf\{t > \alpha_k\colon Z(t, \{0\}) \leq 1 - 1/2\lambda\}$ and $\alpha_{k+1} = \inf\{t > \beta_k\colon Z(t, \{0\}) \geq 1 - 1/m\lambda\}$ are finite for all k. Define $S(t) = 1 - Z(t, \{-1, 0, 1\})$. By Lemma 2.3, there exists a Brownian motion W such that

$$\begin{aligned} S(t) &= S(0) + \int_0^t \sqrt{S(s)(1 - S(s))}\, dW(s) \\ &\quad + \int_0^t \left(\frac{\lambda}{2}Z(s, \{-1, 1\}) - \frac{\lambda}{2}Z(s, \{-2, 2\})\right) ds. \end{aligned} \tag{5.16}$$

Note that on the interval $[\alpha_k, \beta_k]$, $Z(s, \{-1, 1\}) \leq 1/2\lambda$ and $S(\alpha_k) \leq 1/m\lambda$. Hence, by Lemma A4.1, on the interval $[\alpha_k, \beta_k]$, S is bounded by the solution of

$$\hat{S}(t) = \frac{1}{m\lambda} + \int_{\alpha_k}^t \sqrt{\hat{S}(s)\big(1 - \hat{S}(s)\big)}\, dW(s) + \frac{1}{4}(t - \alpha_k). \tag{5.17}$$

Let $\tau = \inf\{t > \alpha_k : \hat{S}(t) = 0\}$. There exist $\delta > 0$ and m_0 such that for $m \geq m_0$, $P\{\beta_k - \alpha_k > \delta\} \geq \delta$. However, for m sufficiently large, $P\{\tau \leq \delta\} > 1 - \delta/2$. Consequently, $P\{\inf_{\alpha_k \leq t \leq \beta_k} S(t) = 0\} \geq \delta/2$, and the conclusion follows by a renewal argument. □

THEOREM 5.12. *Suppose there exist* $A_1, A_2, \dots, B_1, B_2, \dots \subset E$ *with* $B_k \supset A_k$ *such that for* $x \in \bigcup B_k$, $Bf(x) = \lambda(x)\int_E (f(y) - f(x))\eta(x, dy)$, $\lambda_1 \equiv \sup_k \sup_{x \in A_k} \lambda(x)\eta(x, A_k^c) < \infty$, $\lambda_2 \equiv \sup_k \sup_{x \in A_k} \lambda(x)\eta(x, B_k^c) < \frac{1}{2}$ *and* $\lambda_3 \equiv \sup_k \sup_{x \in B_k} \lambda(x)\eta(x, B_k^c) < \infty$. *If for every* n *there exists* $t > n$ *such that* $X_1(t) \in \bigcup A_k$ [*in particular, if* X_1 *is stationary and ergodic and the stationary distribution* π *for the mutation process satisfies* $\pi(\bigcup A_k) > 0$], *then, with probability* 1, *there exist times* t *such that for some* k, $Z(t, B_k) = 1$.

EXAMPLE 5.13. If the mutation process is a finite range random walk on $\mathbb{Z}^d$, then there will be times when $\mathbb{Z}$ is concentrated on a finite set. In particular, if $Bf(k) = \lambda \sum p_m (f(k+m) - f(k))$, then take $A_k = \{k\}$ and $B_k = \{k + m : p_m > 0\}$.

PROOF OF THEOREM 5.12. The finiteness of λ_1 and the recurrence condition imply that for any $\varepsilon > 0$, there will be infinitely many disjoint time intervals $[\alpha_i, \beta_i]$ on which $Z(t, A_k) \geq 1 - (1 - 2\lambda_2)/4\lambda_3$ and $Z(\alpha_i, A_k) \geq 1 - \varepsilon$ for some k. As in the proof Theorem 5.11(c), for ε sufficiently small, $\inf_i P\{\inf_{\alpha_i \leq t \leq \beta_i} Z(t, B_k^c) = 0\} > 0$. □

In sharp contrast to the assertion of Theorem 5.3 that the support of Z is all of E if the mutation operator is bounded and the mutation process is recurrent, Dawson and Hochberg (1982) prove in the case of Brownian mutation that the support of Z is compact. We extend their result to more general continuous mutation processes. Their proof is based on genealogies and is similar to the one presented here.

THEOREM 5.14. *Let* $P(t, x, F)$ *be the transition function for the mutation process and define* $\bar{p}(t, \varepsilon) = \sup_{x \in E} \sup_{s \leq t} P(s, x, B_\varepsilon(x)^c)$. *Let* $C_t = \{x : Z(t, B_\varepsilon(x)) > 0$ *for every* $\varepsilon > 0\}$. *Suppose that there exists a sequence of positive integers* κ_n *and a sequence* $\varepsilon_n > 0$ *such that*

$$\sum_{n=1}^{\infty} \varepsilon_n < \infty, \qquad \sum_{n=1}^{\infty} \kappa_{n+1} \bar{p}\left(\frac{1}{\kappa_n}, \varepsilon_n\right) < \infty, \tag{5.18}$$

Then with probability 1, C_t *is compact.*

REMARK 5.15. (a) In the case of Brownian mutation, $\kappa_n = 2^n$ and $\varepsilon_n = n^{-2}$ satisfy (5.18).

(b) Recall that $\lim_{t \to 0} t^{-1} \sup_x P(t, x, B_\varepsilon(x)^c) = 0$ for each $\varepsilon > 0$ is a classical condition for continuity of the sample paths of a Markov process. See

Dynkin [(1965), Theorem 3.5] and Ethier and Kurtz [(1986), Proposition 4.2.9 and Problem 4.11.17].

(c) The uniform stochastic continuity implied by the assumptions of the theorem will not hold for many diffusions at least under the usual metric. If E is locally compact, however, then for many processes one can show existence of a "stochastic Lyapunov function" V [see, e.g., Kushner (1967)] such that $K_c = \{x: V(x) \le c\}$ is compact for each $c > 0$ and $\bar{p}$ defined by

$$\bar{p}(t, \varepsilon) = \sup_{x \in E} \sup_{s \le t} P(s, x, \{y: V(y) \ge V(x) + \varepsilon\}) \tag{5.19}$$

satisfies the conditions of the theorem. In this case, the argument below shows that $C_t \subset K_c$ for some c.

PROOF OF THEOREM 5.14. Let $\beta_n = \inf\{s: |\Gamma(s,t)| \ge n\}$. [$\Gamma(s,t)$ is as in Section 3.] Then $\lim_{n\to\infty} n(t - \beta_n) = 2$ a.s. The limit follows from the fact that

$$t - \beta_n = \sum_{k=n+1}^{\infty} \frac{2}{k(k-1)} \Delta_k = \frac{2}{n} + \sum_{k=n+1}^{\infty} \frac{2}{k(k-1)} (\Delta_k - 1) \tag{5.20}$$

and the fourth moment of the second term on the right is $O(n^{-6})$.

Let A_{n+1} be the event that some particle with index in $\Gamma(\beta_{3\kappa_{n+1}}, t)$ is a distance greater than ε_n from its ancestor at time $\beta_{3\kappa_n}$. Define $\mathscr{F}_n = \sigma\{A_k, \beta_{3\kappa_k}: k \le n\}$. Then $P(A_{n+1}|\mathscr{F}_n) \le 3\kappa_{n+1}\bar{p}(t - \beta_{3\kappa_n}, \varepsilon_n)$, and hence

$$\sum_{n=1}^{\infty} P(A_{n+1}|\mathscr{F}_n) < \infty. \tag{5.21}$$

By the conditional version of Borel–Cantelli [see, for example, Breiman (1968), page 96], only finitely many of the A_n occur. Let $\delta_n = \sum_{k \ge n} \varepsilon_k$. Then for all n sufficiently large,

$$C_t \subset \bigcup_{j \in \Gamma(\beta_{3\kappa_n}, t)} B_{\delta_n}(X_j(\beta_{3\kappa_n})), \tag{5.22}$$

which implies that C_t is totally bounded and hence compact. □

APPENDIX

A1. Relative compactness for measure-valued processes.

LEMMA A1.1. *Let E be a complete separable metric space and let $\{Z_m\}$ be a sequence of processes with sample paths in $D_{\mathscr{P}(E)}[0,\infty)$. Then $\{Z_m\}$ is relatively compact if and only if for each $\varepsilon > 0$ and $T > 0$ there exists a compact $K \subset E$ such that*

$$P\Big\{\sup_{t \le T} Z_m(t, K^c) \ge \varepsilon\Big\} \le \varepsilon$$

and there exists a linear separating set $D \subset \overline{C}(E)$ such that $\{\langle f, Z_m\rangle\}$ is relatively compact in $D_{\mathbb{R}}[0,\infty)$ for each $f \in D$.

PROOF. Necessity follows from Prohorov's theorem and the continuous mapping theorem. Fix $\eta > 0$, select K_k compact such that

$$P\left\{\sup_{t\le T} Z_m(t, K_k^c) \ge \frac{\eta}{2^k}\right\} \le \frac{\eta}{2^k}$$

and define $\mathscr{K} = \{\mu: \mu(K_k) \ge 1 - \eta/2^k,\ k = 1, 2, \dots\}$. Then $\mathscr{K}$ is compact by Prohorov's theorem and

$$P\{Z_m(t) \in \mathscr{K}, t \le T\} \ge 1 - \eta.$$

The linearity of D ensures that for $f_1, \dots, f_k \in D$, $\{(\langle f_1, Z_m\rangle, \dots, \langle f_k, Z_m\rangle)\}$ is relatively compact in $D_{\mathbb{R}^k}[0,\infty)$ and hence for $g \in C(\mathbb{R}^k)$, $\{g(\langle f_1, Z_m\rangle, \dots, \langle f_k, Z_m\rangle)\}$ is relatively compact. The theorem then follows by Theorem 3.9.1 of Ethier and Kurtz (1986). □

LEMMA A1.2. *Suppose* $\{X_m\}$ *is relatively compact in* $D_E[0,\infty)$, *for each* m, $\{\mathscr{F}_t^m\}$ *is a filtration, and* $Z_m(t)$ *is the conditional distribution of* $X_m(t)$ *given* F_t^m. *Then for each* $\varepsilon > 0$ *and* $T > 0$ *there exists compact* $K \subset E$ *such that*

$$P\left\{\sup_{t\le T} Z_m(t, K^c) \ge \varepsilon\right\} \le \varepsilon.$$

PROOF. For $t \le T$, $Z_m(t, K^c) \le M(t) \equiv E[I_{\{X_m(s)\in K^c,\ \text{some } s\le T\}} | \mathscr{F}_t^m]$. Since M is a martingale, we have

$$P\left\{\sup_{t\le T} Z_m(t, K^c) \ge \varepsilon\right\} \le P\left\{\sup_{t\le T} M(t) \ge \varepsilon\right\} \le \frac{P\{X_m(s) \in K^c, \text{some } s \le T\}}{\varepsilon}$$

and by the relative compactness of $\{X_m\}$ there exists a compact K such that the probability on the right is less than ε^2 for all m. □

A2. Convergence in probability in $D_E[0,\infty)$.

LEMMA A2.1. *Let* $\{X_n\}$ *be a sequence of processes with sample paths in* $D_E[0,\infty)$ *defined on the same sample space. Suppose that* $\{X_n\}$ *is relatively compact in* $D_E[0,\infty)$ (*in the sense of convergence in distribution*) *and that for a dense set* $H \subset [0,\infty)$, $\{X_n(t)\}$ *converges in probability in* E *for each* $t \in H$. *Then* $\{X_n\}$ *converges in probability in* $D_E[0,\infty)$.

REMARK A2.2. The lemma is not correct if convergence in probability is replaced by convergence almost everywhere. Let $\{f_n\}$ be any sequence in $D_{\mathbb{R}}[0,\infty)$ that converges pointwise to zero but does not converge uniformly on bounded intervals, and let $\{\xi_n\}$ be any sequence of random variables, bounded by 1, that converges in probability to zero but not almost surely. Define $X_n = \xi_n f_n$.

PROOF OF LEMMA A2.1. The relative compactness of $\{X_n\}$ implies that $\{(X_n, X_m)\}$ is relatively compact in $D_E[0,\infty) \times D_E[0,\infty)$ (but not necessarily in

$D_{E^2}[0, \infty))$. Let (X, Y) be a limit point as n, $m \to \infty$ of $\{(X_n, X_m)\}$. Then for $t \in H$, the convergence in probability of $\{X_n(t)\}$ ensures that with probability 1, one of the following holds: $X(t) = Y(t)$, $X(t) = Y(t-)$, $X(t-) = Y(t)$, $X(t-) = Y(t-)$. Since H is dense, the right continuity of X and Y then ensures that $X = Y$. Let $d(x, y)$ be a metric giving the Skorohod topology on $D_E[0, \infty)$. Along the subsequence converging to (X, Y), the continuous mapping theorem implies $d(X_n, X_m) \Rightarrow d(X, Y) = 0$. However, this observation applies to any convergent subsequence of $\{(X_n, X_m)\}$ with n and m going to ∞. Consequently, for $\varepsilon > 0$,

$$\lim_{n, m \to \infty} P\{d(X_n, X_m) > \varepsilon\} = 0$$

and convergence in probability of $\{X_n\}$ follows. □

A3. Averages of bounded martingale differences.

Lemma A3.1. *Let $\xi_1, \dots, \xi_n$ be martingale (or reverse martingale) differences with $|\xi_k| \le L$ for some constant L. Then for $\varepsilon > 0$, there exist positive constants C and η depending only on ε and L, such that*

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} \xi_k\right| > \varepsilon\right\} \le Ce^{-\eta n}.$$

Proof. The proof is the same as that of Theorem 5.9 of Kurtz (1972). □

A4. Comparison lemma.

Lemma A4.1. *Suppose X and Y satisfy*

$$X(t) = A(t) + \int_0^t \sigma(X(s))\, dW(s) + \int_0^t b(X(s))\, ds$$

and

$$Y(t) = B(t) + \int_0^t \sigma(Y(s))\, dW(s) + \int_0^t b(Y(s))\, ds,$$

where $V \equiv B - A$ is continuous and nondecreasing and $B(0) > A(0)$. Let

$$\tau_c = \inf\left\{t: \frac{|\sigma(Y(t)) - \sigma(X(t))|}{|Y(t) - X(t)|} + \frac{|b(Y(t)) - b(X(t))|}{|Y(t) - X(t)|} \ge c\right\}$$

and $\tau = \lim_{c\to\infty} \tau_c$. Then with probability 1, $Y(t) > X(t)$ for all $t < \tau$.

PROOF. Observe that

$$\begin{aligned}\ln(Y(t)-X(t)) = \ln(Y(0)-X(0)) &+ \int_0^t \frac{1}{Y(s)-X(s)}\,dV(s)\\ &+\int_0^t \frac{\sigma(Y(s))-\sigma(X(s))}{Y(s)-X(s)}\,dW(s)\\ &-\int_0^t \frac{(\sigma(Y(s))-\sigma(X(s)))^2}{(Y(s)-X(s))^2}\,ds\\ &+\int_0^t \frac{b(Y(s))-b(X(s))}{Y(s)-X(s)}\,ds.\end{aligned}$$

Since the second term on the right is nonnegative, the right side is bounded below on the interval $[0,\tau_c \wedge t]$ for every $c>0$ and $t>0$, implying the conclusion of the lemma. □

A5. Independence of random selections of iid random variables.

LEMMA A5.1. *Let* $(S,\mathscr{S})$ *be a measurable space and let* $\{\xi_m: m=1,2,\dots\}$ *be independent, identically distributed, S-valued random variables with distribution* μ. *Suppose that* $\{A_k, k=1,2,\dots\}$ *are positive integer-valued random variables satisfying* $P\{A_i \neq A_j\}=1$ *for all* $i\neq j$ *and that* $\{A_k\}$ *is independent of* $\{\xi_m\}$. *Then* $\xi_{A_1}, \xi_{A_2},\dots$ *are independent and have distribution* μ, *and* $\{\xi_{A_i}\}$ *is independent of* $\{A_k\}$.

PROOF. It is enough to show that for each m and $h_1,\dots,h_m \in B(S)$,

$$E\left[\prod_{k=1}^n h_k(\xi_{A_k})\middle| A_1,\dots,A_m\right] = \prod_{k=1}^n \int_S h_k\,d\mu,$$

but this identity is a special case of Proposition A.4.5 of Ethier and Kurtz (1986). □

LEMMA A5.2. *Let* $(S,\mathscr{S})$ *be a measurable space and let* $\{\xi_m: m=1,2,\dots\}$ *be exchangeable, S-valued random variables. Suppose that* $\{A_k, k=1,2,\dots\}$ *are positive integer-valued random variables satisfying* $P\{A_i\neq A_j\}=1$ *for all* $i\neq j$ *and that* $\{A_k\}$ *is independent of* $\{\xi_m\}$. *Then* $\{\xi_{A_m}: m=1,2,\dots\}$ *is exchangeable and has the same distribution as* $\{\xi_m: m=1,2,\dots\}$ *and* $\{\xi_{A_m}\}$ *is independent of* $\{A_k\}$.

PROOF. Since $E[g(\xi_{i_1},\dots,\xi_{i_m})]=E[g(\xi_1,\dots,\xi_m)]$ as long as $i_1,\dots,i_m$ are distinct, the independence of $\{A_k\}$ from $\{\xi_m\}$ implies

$$E\left[g(\xi_{i_1},\dots,\xi_{i_m})| A_1,\dots,A_m\right] = E\left[g(\xi_1,\dots,\xi_m)\right]$$

and Proposition A.4.5 of Ethier and Kurtz (1986) gives

$$E\left[g(\xi_{A_1},\dots,\xi_{A_m})|A_1,\dots,A_m\right]=E\left[g(\xi_1,\dots,\xi_m)\right]$$

and the lemma follows. □

Acknowledgments. Much of this research was carried out while the second author was visiting Queen Mary and Westfield College. The hospitality of that institution is gratefully acknowledged.

REFERENCES

BLACKWELL, D. and DUBINS, L. E. (1983). An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.* **89** 691–692.

BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, MA.

CHOW, Y. S. and TEICHER, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales*, 2nd ed. Springer, New York.

DAWSON, D. A. and HOCHBERG, K. J. (1982). Wandering random measures in the Fleming–Viot model. *Ann. Probab.* **10** 554–580.

DONNELLY, P. and KURTZ, T. G. (1996). The asymptotic behavior of an urn model arising in population genetics. *Stochastic Process. Appl.* To appear.

DONNELLY, P. and TAVARÉ, S. (1987a). The population genealogy of the infinitely-many neutral alleles model. *J. Math. Biol.* **25** 381–391.

DONNELLY, P. and TAVARÉ, S. (1987b). A genealogical description of the infinite alleles model. *Stochastic Models in Biology. Lecture Notes in Biomath.* **70** 27–35. Springer, Berlin.

DURRETT, R. (1991). *Probability: Theory and Examples*. Wadsworth and Brooks/Cole, Belmont, CA.

DYNKIN, E. B. (1965). *Markov Processes, I.* Springer, Berlin.

ETHIER, S. N. and GRIFFITHS, R. C. (1987). The infinitely-many-sites model as a measure-valued diffusion. *Ann. Probab.* **15** 515–545.

ETHIER, S. N. and GRIFFITHS, R. C. (1990). The neutral two-locus model as a measure-valued diffusion. *Adv. in Appl. Probab.* **22** 773–786.

ETHIER, S. N. and GRIFFITHS, R. C. (1993). The transition function of a Fleming–Viot process. *Ann. Probab.* **21** 1571–1590.

ETHIER, S. N. and KURTZ, T. G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. in Appl. Probab.* **13** 429–452.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

ETHIER, S. N. and KURTZ, T. G. (1987). The infinitely-many-alleles model with selection as a measure-valued diffusion. *Stochastic Models in Biology. Lecture Notes in Biomath.* **70** 72–86. Springer, Berlin.

ETHIER, S. N. and KURTZ, T. G. (1992). On the stationary distribution of the neutral diffusion model in population genetics. *Ann. Appl. Probab.* **2** 24–35.

ETHIER, S. N. and KURTZ, T. G. (1993). Fleming–Viot processes in population genetics. *SIAM J. Control Optim.* **31** 345–386.

ETHIER, S. N. and KURTZ, T. G. (1995). Coupling and ergodic theorems for Fleming–Viot processes. In preparation.

FLEMING, W. H. and VIOT, M. (1979). Some measure-valued markov processes in population genetics theory. *Indiana Univ. Math. J.* **28** 817–843.

GRIFFITHS, R. C. (1980). Lines of descent in the diffusion approximation of neutral Wright–Fisher models. *Theoret. Population Biol.* **17** 37–50.

HOPPE, F. (1987). The sampling theory of neutral alleles and an urn model in population genetics. *J. Math. Biol.* **25** 123–159.

KINGMAN, J. F. C. (1976). Coherent random walks arising in some genetical models. *Proc. Roy. Soc. London Ser. A* **351** 19–31.

KINGMAN, J. F. C. (1982a). On the genealogy of large populations. *J. Appl. Probab.* **19A** 27–43.

KINGMAN, J. F. C. (1982b). The coalescent. *Stochastic Process. Appl.* **13** 235–248.

KURTZ, T. G. (1972). Inequalities for the law of large numbers. *Ann. Math. Statist.* **43** 1874–1883.

KURTZ, T. G. (1995). Martingale problems for conditional distributions of Markov processes. Preprint.

KURTZ, T. G. and OCONE, D. L. (1988). Unique characterization of conditional distributions in nonlinear filtering. *Ann. Probab.* **16** 80–107.

KUSHNER, H. J. (1967). *Stochastic Stability and Control*. Academic Press, New York.

MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Cambridge Philos. Soc.* **54** 60–71.

MORAN, P. A. P. (1975). Wandering distributions and the electrophoretic profile. *Theoret. Population Biol.* **8** 318–330.

MORAN, P. A. P. (1976). Wandering distributions and the electrophoretic profile. II. *Theoret. Population Biol.* **10** 145–149.

OHTA, T. and KIMURA, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetics Res.* **22** 201–204.

OHTA, T. and KIMURA, M. (1974). Simulation studies on electrophoretically detectable genetic variability in a finite population. *Genetics* **16** 97–159.

OVERBECK, L., RÖCKNER, M. and SCHMULAND, B. (1995). An analytic approach to Fleming–Viot processes with interactive selection. *Ann. Probab.* **23** 1–36.

SCHMULAND, B. (1991). A result on the infinitely many neutral alleles diffusion model. *J. Appl. Probab.* **28** 253–267.

SHIGA, T. (1982). Wandering phenomena in infinite allelic diffusion models. *Adv. in Appl. Probab.* **14** 457–483.

SVERCHKOV, M. and SMIRNOV, M. (1990). Maximal coupling of D-valued processes. *Sov. Math. Dokl.* **41** 352–354.

SCHOOL OF MATHEMATICAL SCIENCES
QUEEN MARY AND WESTFIELD COLLEGE
MILE END ROAD
LONDON E1 4NS
UNITED KINGDOM
E-MAIL: p.j.donnelly@qmw.ac.uk

DEPARTMENTS OF MATHEMATICS AND STATISTICS
UNIVERSITY OF WISCONSIN–MADISON
VAN VLECK HALL
480 LINCOLN DRIVE
MADISON, WISCONSIN 53706-1388
E-MAIL: kurtz@math.wisc.edu